

PCT

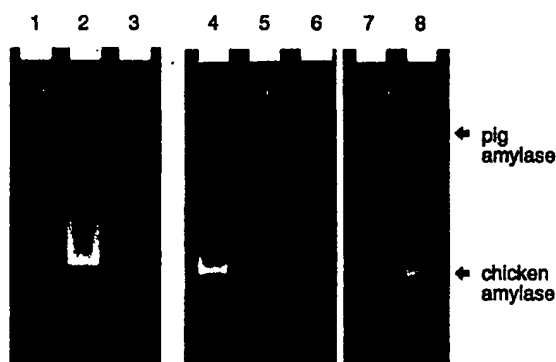
WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|--|--|--|--|
| (51) International Patent Classification ⁶ : C12N 15/00, A01K 67/027, C12N 15/86, 9/72 | | A1 | (11) International Publication Number: WO 99/43795 |
| | | | (43) International Publication Date: 2 September 1999 (02.09.99) |
| (21) International Application Number: PCT/CA98/00607 | | (74) Agents: ERRATT, Judy, A. et al.; Gowling, Strathy & Henderson, Suite 2600, 160 Elgin Street, Ottawa, Ontario K1P 1C3 (CA). | |
| (22) International Filing Date: 25 June 1998 (25.06.98) | | | |
| (30) Priority Data: 2,224,108 24 February 1998 (24.02.98) CA | | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). | |
| (71) Applicant (for all designated States except US): HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF AGRICULTURE AND AGRI-FOOD CANADA [CA/CA]; Building 34, Centre for Food and Animal Research, Central Experimental Farm, Ottawa, Ontario K1A 0C6 (CA). | | Published With international search report. | |
| (72) Inventors; and (75) Inventors/Applicants (for US only): GAVORA, Jan, S. [CA/CA]; 51 Forest Hill Avenue, Ottawa, Ontario K2C 1P7 (CA). FALCONER, Marcia, M. [CA/CA]; 33 Abingdon Drive, Nepean, Ontario K2H 7M5 (CA). NGUYEN, Thuy, H. [CA/CA]; Apartment 1201, 1316 Carling Avenue, Ottawa, Ontario K1Z 7L1 (CA). BENKEL, Bernhard, F. [CA/CA]; 293 Summit Avenue, Ottawa, Ontario K1H 5Z7 (CA). | | | |

(54) Title: TRANS-SOMATICS WITH GENE TRANSFER INTO MAMMARY EPITHELIAL CELLS



(57) Abstract

A method is described to transfer a gene encoding a valuable compound, such as a pharmaceutical, into the secretory cells of the mammary gland to produce a new compound into the milk or to alter the composition of the milk. In this method the packaging cell line producing the viral-derived particles is infused into the mammary gland. The packaging cells will attach and survive for a period of time within the mammary gland. While the cells are viable, they will supply a continuous source of viral-derived particles to trans-infect the maximum number of mammary epithelial cells. After a period of time in the mammary gland, both the particles and the packaging cells will be destroyed by natural mechanisms while the trans-infected mammary epithelial cells continue to express gene(s) encoding the valuable compound or gene(s) to alter the composition of the milk. One or more genes can be trans-infected including DNA sequences that contribute to the efficient production of an active compound or to its stability. The packaging cells and the viral-derived particles used in this method can be those which trans-infect dividing or non-dividing cells and can be used either singly or together. This method can be used alone or in combination with other novel methods designed to ensure that the viral-derived particles are correctly positioned to trans-infect the mammary epithelial cells. Increasing the trans-infection of the mammary epithelial cells with the viral particles will result in a higher concentration of the valuable compound in the milk.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | R | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

TRANS-SOMATICS WITH GENE TRANSFER INTO MAMMARY EPITHELIAL CELLS

5

FIELD OF INVENTION

The present invention relates to a method of producing value-added milk by the incorporation of specific DNA sequence(s) into the cells of the mammary gland. The term "value-added milk" is meant to mean milk containing a valuable compound, such as a pharmaceutical, as well as milk with a changed composition so that the market value of the milk is enhanced.

10

BACKGROUND OF THE INVENTION

This invention is based upon a technique to transfer a gene, and related non-translated control sequences, into the secretory cells of the mammary gland to produce new compounds in milk and/or to alter the milk composition.

15

An example of a compound produced into milk by this method would be a pharmaceutical which cannot be manufactured in a biologically active state. This invention provides an alternative method of producing such pharmaceuticals into milk using the synthetic capabilities of the mammary epithelial cell.

20

Mammals produce copious amounts of complex proteins into milk to provide nutrition for their young. To harness this capability, the DNA of, for example a valuable pharmaceutical, can be inserted into a mammary epithelial cell and this cell then will produce the active, pharmaceutical compound into the secreted milk. In a cow, the udder will act as a receptacle to hold the milk until it can be collected. The milk containing the added compound can be processed to extract and purify this pharmaceutical compound for subsequent sale, possibly to the medical and/or veterinary communities, or the milk could be consumed directly as a therapeutic

25

30

agent.

At present there are two basic techniques that can be used to produce value-added milk. The first is to create a transgenic animal by microinjection or transfection of foreign DNA into an ovum or a fertilized egg. Incorporation of DNA at this stage in development generally results in a **transgenic** animal which carries the inserted DNA in every cell.

There are positive and negative aspects of producing a transgenic animal which expresses foreign proteins in its milk. A positive point is that a single founder animal can create a population of transgenic animals by natural reproduction. However, progenies do not always produce the exogenous protein at the same level as the original animal. Negative aspects include the technically difficult procedures required to produce the animal and the long time between adding the foreign DNA and harvesting the exogenous protein. In addition, the presence of even very small amounts of an active pharmaceutical in every tissue may be detrimental to the health of the animal.

An alternative method is to produce value-added milk by adding the desired DNA only to the cells of the mammary gland of the animal. This results in a **trans-somatic** animal (or chimera) which contains the inserted DNA essentially in only one tissue, the mammary gland.

Trans-somatic animals have the advantage that they can be produced with less technical difficulty. They also can be produced quickly so that there is a period of only weeks to months between adding the foreign DNA and harvesting the exogenous protein. Moreover, since only one tissue contains the added DNA and produces the resulting compound, health risks to the trans-somatic animal are reduced. Although the DNA is not passed on to the progeny, this is compensated for by the ease and speed with which a trans-somatic animal can be produced.

A trans-somatic goat which expresses human growth hormone (hGH) into milk has been produced by Archer et al. (1994). In Archer the viral-derived particles were infused into the mammary gland for approximately every two days for two weeks. The levels of the compound, human growth hormone, which was used as an example, were very low and approached background levels after the first day. Also Gould et al. (United States Patent 5,215,904) described a method for increasing the rate of mitosis of mammary epithelial cells and then exposing these cells to viral particles for integration of the desired DNA into the epithelial cell.

To produce either a trans-somatic or a transgenic animal, exogenous DNA must pass through the exterior cell membrane. Eukaryotic cells have evolved a membrane which is impervious to most substances including heterologous DNA. Numerous techniques have been developed to bypass this barrier. These include:

- electroporation,
- carrier lipids (liposomes, negative, positive or neutral charged vesicles),
- mechanical wounding of cells including microinjection, liquid or air-jet pressure and scrape loading,
- use of particles composed partially or wholly of viral proteins.

Methods which have successfully produced trans-somatic animals include:

1) Arterial injection of DNA carried in liposomes (small lipid vesicles) was used to produce a trans-somatic mouse (Thierry et al. 1995). This technique can be adapted to deliver DNA to the mammary gland by injection into the major artery serving the mammary gland but circulation of the blood carrying the DNA can result in transfection of multiple tissues.

2) Direct injection of the DNA into tissues was used successfully to add foreign DNA to muscle and other tissues (Furth et al. 1992). This technique can be adapted to inject virus-like particles, carrying the foreign DNA, directly into the tissue of the udder.

3) Use of viral-derived particles carrying DNA coding for human growth

hormone (hGH) were infused through the teat canal, for example by Archer et al. (1994). This resulted in production of trans-somatic goats which expressed hGH into the milk.

5 Viruses reproduce within cells and therefore have evolved a technique to bypass the protective cell membrane to deliver the viral genome (DNA) into a host cell. To enter a cell, protein(s) of the outer viral shells first bind to receptors on the cell surface and then the virus is internalized.

10 The method used by Archer involves transfecting a cell line with DNA coding for various, but not all, proteins of a virus. This cell line, called a “**packaging cell line**”, will produce empty virus shells which can bind to receptors on the host cell membrane. When heterologous DNA, coding for a pharmaceutical or other milk modification, is transfected into the packaging cell line, this DNA will be packaged
15 into the viral-derived particle. When the viral-derived particle comes in contact with a milk-producing cell, the viral proteins of the shell ensure that the heterologous DNA is carried into the cell. Other viral proteins, associated with the particle, integrate the heterologous DNA into the genome of the host cell so that the protein encoded by the DNA can be expressed. In this method the viral-derived particles are used to
20 introduce the heterologous DNA into the mammary gland.

 The trans-somatic methods of the prior art offer advantages over the transgenic method; however the very low levels of foreign protein in the milk of the trans-somatic animal have limited the commercial success of these methods.

25

 Thus the present invention is directed to methods of improving the yield of the foreign protein in the milk of a trans-somatic animal.

SUMMARY OF THE INVENTION

30 The present invention relates to a method of producing valuable compounds into milk and/or changing the composition of milk so as to enhance its properties

and/or its marketability. More specifically the present invention relates to the addition of specific DNA sequences, including non-translated regulatory sequences, to the cells of the mammary gland and the subsequent expression of compound(s) encoded by that DNA into the milk. Other DNA sequences that enhance the efficiency of production of the compound, enhance the stability of the compound, or result in biological activity of the compound can also be added to the mammary epithelial cells either at the same time or at a different time.

The novel method of the present invention involves the use of viral-derived particles and packaging cells which produce these particles for infusion into the mammary gland through the teat canal. The packaging cells will attach and survive for a period of time within the mammary gland. While the cells are viable, they will supply a continuous source of viral-derived particles. These viral-derived particles trans-infect only dividing cells and are destroyed relatively quickly in the mammary gland. Thus a continuous supply of viral-derived particles from the packaging cells present in the mammary gland will ensure that viral-derived particles are present and can trans-infect the mammary epithelial cells whenever they divide.

The packaging cells and viral-derived particles can be from retroviruses and from non-retroviruses. Most retroviral-derived particles trans-infect only dividing cells. Non-retroviral particles such as those from adenovirus, Epstein-Barr virus, or other viruses trans-infect non-dividing cells. Thus a mixture of the two types of particles, and the associated packaging cells if needed, will ensure delivery of the packaged DNA to the maximum number of cells. Moreover, DNA can be packaged into viral-derived particles *in vitro* and these can be used for trans-infection as well.

This method can be used alone or in combination with other novel methods designed to ensure that the viral-derived particles are correctly positioned to trans-infect the mammary epithelial cells. Increasing the trans-infection of the mammary epithelial cells with the viral particles will result in a higher concentration of the valuable compound in the milk along with possible other compounds produced from

the inserted DNA that will enhance the efficient production, stability or activity of the compound.

Thus according to the present invention there is provided a method of
5 producing a trans-somatic mammal, wherein said method provides the incorporation
of a DNA sequence into the secretory cells of the mammary gland to alter the
composition of the milk, comprising the steps of: providing a vector containing a
DNA sequence encoding a valuable compound; packaging said vector into a cell line;
preparing a solution comprising the packaged vector and cell line producing said
10 packaged vector; and delivering said solution into the mammary gland to allow the
incorporation of the DNA into the secretory cells of the mammary gland.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features of the invention will become more apparent from the
15 following description in which reference is made to the appended drawings wherein:
FIGURE 1 shows the presence of amylase in an udder infused with Clone 10 (left
hind quarter) and Clone 12 (right hind quarter). The left front quarter was left
untouched as a negative control and the right front quarter was infused with DMEM
and Polybrene but no cells, to serve as a further negative control. Lanes 1, 2 & 3, early
20 premilk from 3 quarters of cow #99. Lane 1, control quarter (RF-); Lane 2, treated
quarter (RH+); Lane 3, treated quarter (LH+). Lanes 4, 5 & 6, late premilk from cow
#99. Lane 4, treated quarter (LH+); Lane 5, treated quarter (RH+); Lane 6, control
quarter (RH-). Lane 7, mixture of pig and chicken amylase standards; Lane 8, chicken
amylase standard.

25 **FIGURE 2** is a Western blot showing the presence of tPA in an udder infused with
Clone 1. Lane 1, premilk from treated quarter (LH+) of cow #56; Lane 2, premilk
from control quarter (LF-) of cow #56; Lane 3, premilk from treated quarter (RF+) of
cow #90.

30 DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention relates to a method to transfer a gene or genes, and

related non-translated control sequences into the secretory cells of a mammary gland to produce "value-added milk". The term "value-added milk" is meant to mean milk containing a valuable compound, such as a pharmaceutical, as well as milk with a changed composition so that the market value of the milk is enhanced.

5

Pharmaceuticals according to the present invention can include, but are not limited to: a tissue plasminogen activator, a blood clotting factor, an antibody, a protein to aid in weight reduction, a galactosyltransferase, a growth factor, an oncoprotein, a protease inhibitor, a hormone, a milk protein, a hormone receptor, a tumor suppressor protein, an aging inhibitor, or an erythropoietin.

10

According to the present invention there is provided a suitable DNA vector (plasmid) containing the desired DNA sequence encoding the heterologous protein (valuable compound) and associated regulatory sequences such as promoters, enhancers, introns, signal sequences, etc. Other sequences to produce compounds that increase efficient production, enhanced stability or are involved in biological activity of the compound can be included in the same or another DNA vector.

15

In one embodiment the vector is based on the pLXSN, pLNCX, or pLNSX plasmids provided under license from Fred Hutchinson Cancer Research Center.

20

According to the present invention the delivery of the DNA to the cells of the mammary gland is accomplished by the use of viral-derived particles and packaging cells, which produce these particles, for infusion into the mammary gland through the teat canal. This infusion is a standard veterinary practice and usually involves the use of a cannula for insertion into the teat canal, the cannula being attached to a syringe containing a solution of the packaging cells and viral-derived particles.

25

It has already been established that such particles derived from retrovirus trans-infect only dividing cells. Thus, the infusion of the particles and the packaging cells takes place when mammary cell division is naturally at a high level. In the

30

present invention heifers at 5 to 7 months pregnancy were used. It would also be possible to use non-pregnant mammals treated with hormones to induce mammary cell division and lactation prior to particle and cell infusion. However, this method would involve additional handling of the animal and thus is not preferred.

5

Other types of viral-derived particles and associated packaging cell lines, such as those based on adenovirus, Epstein-Barr virus, or other viruses can also be used. These non-retroviral derived particles can infect non-dividing cells. If used in combination with the retroviral derived particles described above, the maximum number of cells will receive the desired DNA. In addition, particles to deliver DNA can be manufactured *in vitro* and use of these alone or in combination with the above described particles will enhance the number of mammary epithelial cells which incorporate the desired DNA. An example of particles which are constructed *in vitro* is described or referred to in Morsy and Caskey (1997).

15

The present invention is applicable to all mammals and is especially applicable to all non-human mammals. Goats, sheep and cows are preferred. Cows with their inherent large volume milk production are particularly preferred.

20

The packaging cells will attach and survive for a period of time within the mammary gland. The reason for the presence of the cells is to supply a continuous source of the viral-derived particles to trans-infect the maximum number of mammary epithelial cells. Both the packaging cells and the viral-derived particles are eventually destroyed in the mammary gland; however, considering the fragile nature of the particles *in vitro*, they are likely destroyed within days. The cells, by contrast, may persist much longer. When cells were infused into the udder, and the udder subsequently flushed and assayed for the presence of cells (dead and alive), most of the cells could not be flushed out. Cells were seen in the wash for 3 days after infusion; of those cells that were flushed out, a large majority remain viable, suggesting that most of the cells have become attached to the udder. In prior art methods (for example Archer et al. 1994), the particles are infused into the udder

25

30

approximately every two days for two weeks. In the present invention no subsequent infusions are necessary. Thus an advantage of the present invention over the prior art is a reduction in the handling of the animal.

5 In prior arts methods (for example, Archer et al. 1994), large scale tissue culture was required to supply sufficient numbers of particles for multiple infusions. An advantage of the present invention over the prior art is that preparation of a large number of particles is not required.

10 In prior arts methods (for example, Archer et al. 1994), the preparation of a stock solution of viral-particles required ultra-centrifugation and resuspension of the particles. A further advantage of the present invention is that no ultracentrifugation is involved. Centrifugation and other handling techniques can destroy the relatively
15 fragile particles. In the present invention, handling of the particles is minimal, increasing the probability that the particles present will remain intact and useful.

The above cited advantages are in addition to the improvement in yield of the compound in the milk, which has already been described above.

20 This method can be used alone or in combination with other methods of the present invention to increase the incorporation of DNA from the viral-derived particles. The additional methods of the present invention are all designed to ensure that the viral-derived particles are correctly positioned to trans-infect the mammary epithelial cells.

25 In one embodiment, the teat canal and udder is first emptied by milking or under negative pressure. Then the udder is flushed with an osmotically-balanced solution which is infused into the teat canal and udder until the udder is full. This solution is removed by milking or under negative pressure and the desired suspension
30 is infused into the udder. In an alternative procedure, the teat canal is not emptied first but is directly filled with an osmotically-balanced solution which is then removed

by milking or under negative pressure.

5 Flushing of the udder serves two purposes; it removes the thick, secreted fluid that is normally present and it forces open the ductwork to allow better access of the packaging cells and viral-derived particles which will be inserted in the next step of the procedure. An example of a suitable osmotically-balanced solution includes saline solution, but can include any other buffered solutions, and also can include the packaging cell grown medium.

10 External massage applied several times a day to the mammary gland improves the circulation of the viral-derived particles and results in more secretory cells in the mammary gland being exposed to the viral-derived particles and thus more cells will incorporate the DNA carried in these particles.

15 In another embodiment, the packaging cells are grown on a commercially available matrix (designed to support growth and replication of tissue culture cells). This solid support matrix can include gelatin, glass, collagen or plastic beads. Cytodex beads or Cultisphere (purchased from Sigma) are two specific examples of useful support means. The beads, with cells adhering to them, are infused into the mammary gland through the teat canal along with a suspension of DNA-containing viral-derived particles. The beads remain in the mammary gland. External massage applied several times a day to the mammary gland recirculates the beads and helps to distribute the viral-derived particles that are produced by the packaging cells growing on the beads. The presence of the packaging cells on beads also ensures the continuous presence of many more DNA-containing viral-derived particles. Concomitantly more secretory cells in the mammary gland will incorporate the DNA carried within these viral particles and the production of the desired protein will increase accordingly. The beads and any cells remaining on them will eventually be removed when milking is begun.

30

In another embodiment, the infusion of cells (with or without growth on

beads) and viral-derived particles containing the desired DNA is followed by infusion of a substance which is more dense than the water-based suspension fluids, used for cell infusion, such as growth medium (Dulbecco's modified Eagle's medium [DMEM], phosphate buffered saline [PBS], etc.). This dense fluid, which in one example is composed of silicone, displaces the aqueous solutions containing the packaging cells and the viral-derived particles upwards into the ductwork of the mammary gland. This prevents collection of the packaging cells in the cistern of the udder and positions the viral-derived cells and the viral-derived particles up into the ducts of the mammary gland. Thus, the viral-derived particles are positioned near the dividing cells in the alveoli of the mammary gland and allow more of the DNA carried in the viral-derived particles to be incorporated. Any physiologically compatible inert fluid that has a density greater than that of the infusion solution can be used according to this embodiment of the present invention.

Thus, the present invention consists of the following procedure, which can be used alone or in combination with optional methods of the present invention, to deliver heterologous DNA to milk producing cells. In this procedure the following steps occur:

- 1) A vector containing the desired DNA sequence(s) is constructed and is transfected by standard means into a packaging cell line.
- 2) A solution containing the packaging cell line, producing viral-derived particles containing the desired DNA sequence(s), and viral-derived particles, is infused into the mammary gland through the teat canal. The packaging cells attach to the epithelial cells of the mammary gland, remain viable and produce viral-derived particles.
- 3) The DNA becomes incorporated into the secretory cells of the mammary gland.
- 4) The milk containing the product induced by addition of the desired DNA is milked from the cistern. The product is purified from the milk or the milk, containing the product, is consumed.

The basic technique can be modified by the addition of one or more of the following steps:

- 1) The udder is flushed with an osmotically-balanced solution.
- 2) The packaging cells are grown on a matrix designed to support proliferation of eukaryotic cells, such as Cytodex beads, and then infused through the teat canal. The cells will remain in the mammary gland and continue to produce viral-derived particles for a period of at least 3 days.
- 3) After infusion of the packaging cells (either as a suspension or grown on a matrix), a compound which is denser than aqueous solutions, such as silicone, is infused into the mammary gland to force the aqueous solutions carrying the packaging cells and viral-derived particles up and into the region of the mammary gland where the DNA can be incorporated into the milk producing cells. The dense compound, if it has been added, and any unattached cells as well as the matrix, if used, is removed from the mammary gland cistern by milking at an appropriate time after infusion. An appropriate time is defined as after the majority of packaging cells attach and after there is sufficient production of viral-derived particles. An example of an appropriate time would be at least 3 days, however shorter or longer periods may also be used. The packaging cells which are not removed at this time die and are removed by the recipient's natural mechanisms.
- 4) External massage applied several times a day to the mammary gland will recirculate the viral-derived particles and the packaging cells, provided either as a suspension or grown on a matrix. This helps to distribute the viral-particles and the cells and increases the incorporation of the DNA carried in these particles into the secretory cells.

While this invention is described in detail with particular reference to preferred embodiments thereof, said embodiments are offered to illustrate but do not limit the invention.

EXAMPLES

Example 1: Preparation of plasmids carrying the desired gene and various control sequences for use in retroviral particle delivery to bovine mammary epithelial cells

5 As a system to transfect the mammary epithelial cells with a desired gene, the PG13 packaging cell line was acquired under license from Fred Hutchinson Cancer Research Center. This system was used because it produces retroviral particles containing the gibbon ape leukemia virus envelope (Galv) which facilitates transfection of bovine cells. To package the desired gene "X" into retroviral particles,
10 the plasmids, pLXSN, pLNCX, pLNSX and pLN, also were acquired under license from Fred Hutchinson Cancer Research Center.

Modifications of plasmids:

15 In order to have an alternative method to select cell clones containing the desired gene, we replaced the *neomycin* (*neo*) gene with the *hygromycin* (*hyg*) gene in pLXSN, pLNSX and pLNCX. This was done by long-range PCR-amplification of the region around the *neo* gene then ligating the PCR product with the *hyg* gene which was amplified from the plasmid pREP4 (purchased from Invitrogen). The resulting plasmids were called pLXSH, pLHSX and pLHCX where "L" represents the Moloney
20 murine virus long terminal repeat (LTR) acting as a promoter, "S" is the SV40 promoter and "C" is the cytomegalovirus promoter region.

25 To put the desired gene, "X", under the control of a constitutive promoter, the cytomegalovirus promoter region in pLNCX and pLHCX was removed by restriction digestion and replaced with the beta actin promoter to make pLNAX and pLHAX. The beta actin promoter sequence was derived using PCR from the pJ6 Ω plasmid purchased from ATCC (catalog no. 37723).

30 In order to avoid transcription interference between the gene used for selection and the gene used to produce the desired protein, transcription from both genes was coupled by replacing the SV40 promoter from the plasmids pLXSN and pLXSH with

an internal ribosomal entry site (IRES). The resulting plasmids, pLAI_N and pLAI_H, express both the selection gene and the inserted gene under the same (LTR) promoter. The IRES in these plasmids provides the translation initiation site within the transcript, allowing the downstream gene product to be produced. The IRES used
5 here is identical to the IRES found in the plasmid pIRES1_{neo} purchased from Clontech. The sequences for pLAI_N and pLAI_H, where A in this case is chicken amylase but which can be any desired protein, are shown in SEQ ID No:1 and SEQ ID No: 2, respectively.

10 In order to increase retroviral titre or stability of the transcript, or to increase the expression level of the desired gene during lactation, or to allow translation of more than one protein from the same transcript, the basic plasmid, pLNCX, was modified. In one modification, the selection gene was removed to minimize the size of the resulting plasmid, pLX. In another modification, the CMV promoter was
15 replaced with the murine mammary tumour virus (M) LTR promoter to improve transcription of the resulting plasmid, pLNM_X, during lactation. The sequence for pLNM_X is shown in SEQ ID No: 3.

In another modification, a wild-type IRES was modified so that the ATG
20 codon at position 10 is destroyed and the sequence downstream of the ATG codon at position 11 codes for the desired gene, "X", in a plasmid such as pLNM_i₂X. The sequence for this IRES modification (*i*₂) is included within the sequence for pLNM_i₂X, shown in SEQ ID No: 4.

25 **Example 2: Preparation of cell clones producing viral-derived particles carrying the chicken amylase gene as a "marker" protein**

The two packaging cell lines used in this experiment were purchased from ATCC, PA317 (catalog no. CRL-9078) and PG13 (catalog no. CRL-10686). A description of both PA317 and PG13 and their use can be found in Miller et al. 1990.

30

To develop and optimize the transfection of mammary epithelial cells using

the retroviral particle delivery system, we used a stable and readily detectable marker protein, chicken amylase, in some experiments. Chicken amylase migrates at a unique position by native gel electrophoresis and can be differentiated from bovine or other amylases. To produce a PG13-derived cell line (a clone) producing retroviral particles which carried chicken amylase under the control of the beta actin promoter, the following procedure was done.

The pLH(A)amy plasmid with *hygromycin* (H) driven by the LTR promoter and with *amylase* (amy) driven by the beta actin promoter (A), was produced by standard recombinant techniques. pLH(A)amy was transiently transfected by the calcium phosphate technique into the PA317 packaging cell line. The transfected PA317 cells produce viral particles containing the amy RNA into the supernatant. The amphotropic viral envelope protein of these particles allows entry into cells of most species including the PG13 packaging cells in a process called trans-infection. Empirically it has been determined that trans-infection produces PG13 clones with a higher rate of particle production as compared to PG13 clones produced by other means of DNA insertion (particle bombardment, calcium phosphate or liposome transfection). Therefore, the viral-derived particles in the supernatant of the transfected PA317 cells were used to trans-infect the PG13 packaging cell line. The resulting clones were selected using 700 ug/ml hygromycin in DMEM with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin for 14 days. Clones were picked using cloning rings and grown as separate cell lines.

These PG13 amy clones were tested for amylase production by native gel electrophoresis of α -amylase. The α -amylase samples in loading buffer (0.1M tris-borate pH 8.5, 10% sucrose, 10mM magnesium chloride) were electrophoresed for 3-4 hours at 250 volts, with cooling, on a 5.5% acrylamide gel containing 0.1M tris-borate pH 8.5. The electrophoresis buffer was 0.1M tris-borate pH 8.5. The gel was then placed with gentle shaking for 1 h at room temperature in a solution of 2% soluble starch, 10mM calcium chloride, 50mM tris-HCl pH 7.5. The gel was briefly rinsed in water, then placed in a dilute solution of iodine (I_2 : KI : water is about 1 : 2 :

2000) until the amylase signal shows as a clear band on the gel.

5 All packaging cells produce filled retroviral particles which contain the desired gene and empty particles which do not carry the desired gene. To increase the number of filled retroviral particles, hygromycin-resistant PG13 amy clones producing high amounts of amylase were **supertransfected** with a second plasmid, pLN(A)amy, which also carried the amylase gene but which has *neomycin* as the clone selection agent. In this procedure, first pLN(A)amy was transfected into the PA317 cell line by the calcium phosphate procedure, and then the viral-derived particles in the
10 supernatant of the transfected PA317 amy-containing cells were used to trans-infect a PG13/pLH(A)amy clone. The PG13/pLH(A) clone which was supertransfected was chosen on the basis of high amylase production. Superclones, containing both pLN(A)amy and pLH(A)amy sequences were obtained by selecting in the presence of 1000 ug/ml G418 and 700 ug/ml hygromycin as described above. The surviving
15 clones were picked and grown up for further testing. Although a marker protein, amylase, was used in these experiments to optimize filling of particles, it is obvious that the same procedure can be done where another protein, such as a pharmaceutical protein, is used instead of the amylase.

20 The superclones producing the highest levels of amylase were analyzed for particle production by the following method. For each clone to be tested, the supernatant containing viral-derived particles was used to trans-infect HeLa 229 (human cervical carcinoma) (purchased from ATCC catalog no. CCL-2.1) and Et2 (bovine mammary) cells (provided by Dr. Boris Zavizion, University of Vermont,
25 Burlington, Vt.). Depending on the resistance gene contained in the plasmid, the trans-infected cells were treated with either 700 ug/ml hygromycin or 1000 ug/ml neomycin, or both, and the resulting colonies were counted. Each colony is the result of one infective (filled) particle. The clone producing the highest number of colonies was selected to be grown for subsequent infusion into the heifer udder. In the specific
30 example shown in Figure 1, clone 10 produced the highest number of viral-derived particles and was designated as PG13/LH/LN(A)amy.

Before a clone can be infused into the udder, it must be tested to determine that it does not produce replication-competent (called "helper") virus. Gene therapy is based on the assumption that the viral-derived particles can carry the desired DNA (i.e. cDNA for amylase or another protein) into the target cell but that once in the cell, the virus cannot reproduce itself. There are two methods to check if helper virus is being produced by the trans-infected packaging cell line. These are described in detail in Cepko, 1992. The first method is to analyze the supernatant of the HeLa and Et2 cells for horizontal spread of hygromycin and/or neomycin resistance. The second method is to look for the presence of reverse transcriptase above background levels in the supernatant of the trans-infected HeLa and Et2 cells. If trans-infection with the supernatant from HeLa and Et2 cells does not produce hygromycin and/or neomycin resistant colonies and if there is no reverse transcriptase above the level found in control cells, then it can be assumed that the particle producing clone, in this case PG13/LH/LN(A)amy, is not producing "helper" virus and therefore can be used for infusion into the udder. We followed the procedures detailed in the Cepko reference and found neither colony nor reverse transcriptase production.

Example 3: Infusion of clones producing retroviral-derived particles which carry a desired gene such as the amylase marker protein

In one example, the cell clone PG13/LH/LN(A)amy, which expresses chicken amylase as a marker protein, was infused into the udder of a 7 month pregnant heifer, #99. Before infusion with the cell clone, each quarter of the udder was flushed with 250ml of saline solution at 37° C using the catheter portion of an angiocath G-18 catheter attached to a 140 cc Monoject syringe with a Luer lock attachment (both purchased from CDMV, Saint-Hyacinthe, Quebec). The infusate was then milked out and measured to determine the volume of cell solution which could be administered to this particular quarter of this specific heifer.

Two superclones, designated as PG13/LH/LN(A)amy Clone 10 and PG13/LH/LN(A)amy Clone 12, were grown in standard tissue culture flasks until the total cell number (in an appropriate number of flasks) was $>2 \times 10^8$ cells/clone. The

cells were trypsinized, counted and resuspended at 1×10^8 cells in 250 ml DMEM without serum but containing 80 ug/ml of Polybrene to facilitate particle adsorption to the mammary epithelial cells.

5 The udder has four separate quarters. The left front quarter was left untouched as a negative control (LF-). The front right quarter was flushed with saline and infused with 250 ml DMEM plus 80 ug/ml Polybrene but with NO cells added and served as a negative control (RF-). The right hind quarter was infused with 1×10^8 Clone 12 cells in 250 ml DMEM plus Polybrene (RH+) and the left hind quarter was
10 infused 1×10^8 Clone 10 cells in 250 ml DMEM plus Polybrene (LH+). The infusate was not removed. The udder was massaged 3 times a day, morning, noon and night, to help to distribute the cells which tend to settle into the cistern of the udder. Three weeks post infusion, the udder was "stripped" (all secretions were milked by hand from each quarter) which removed about 7 to 10 ml per quarter of a viscous, serum-
15 like fluid. This was designated as "early premilk". Eight weeks post infusion the udder was stripped again, producing about 25 ml per quarter of a slightly cloudy, viscous fluid designated as "late premilk". The samples from the right front control quarter RF(-) and the two infused quarters, the right hind, RH(+) and the left hind, LH(+) were analyzed for the presence of active amylase as previously described (see
20 Figure 1).

Figure 1 shows that in the early premilk sample, amylase is present in both infused quarters (RH) and (LH) while the control quarter (RF) contains no amylase. In the late premilk, the negative control quarter (RF) continues to show no amylase
25 activity. Clone 10, left hind quarter, shows the highest amylase activity. Clone 12, right hind quarter, has decreased activity compared to the early premilk sample, but a small amount of amylase activity was detected. The samples loaded represent equal volumes (100 ul/lane) of the premilk samples. No attempt was made to load equal amounts of protein. The apparent reduction in the amount of amylase present between
30 early and late premilk samples is caused by increased volume of the late premilk samples relative to the volume of the early premilk samples.

Example 4: Preparation of a clone producing viral-derived particles that carry the tPA gene and infusion of this clone into the udder of a pregnant heifer

Tissue plasminogen activator (tPA) is used to treat heart attack and stroke by dissolving blood clots. We selected human tPA to be the first pharmaceutical product to be produced by our method in bovine milk. The ptPA-K plasmid, containing the mutated cDNA sequence for human tPA (where the amino acids KHRH 296-299 was mutated to AAAA, i.e. the "K" mutation), was purchased from ATCC (American Type Culture Collection, catalog no. 68059). Two additional mutations were incorporated into ptPA-K to form ptPA-TNK: the amino acid threonine at position 103 was mutated to asparagine (the "T" mutation), and the amino acid asparagine at position 117 was mutated to glutamine (the "N" mutation). The tPA amino acid sequence and a description of the above modifications can be found in Pennica et al. 1983. Both mutations were produced by using mismatched oligonucleotides containing the altered nucleotide sequence as primers for PCR amplification. The tPA-TNK gene was subsequently excised by restriction digestion and ligated into the pLXSH plasmid to make pL(tPA)SH.

The resulting PG13(tPA) clones were analyzed for tPA production by colorimetric determination using Spectrozyme (# 444 purchased from American Diagnostica Inc.). Clones which showed high levels of tPA were then checked for production of filled particles by colony counts (as described previously) and were safety checked to insure that no replication competent virus was being produced and that there was no reverse transcriptase production above that found as background in HeLa or ET2 cells. None of the clones tested produced replication competent virus by either of these tests. The clones with the highest particle production and highest level of tPA production were selected to be grown for infusion into the udder. In the following example, clone PG13/L(tPA)SH-1, was chosen for infusion into the udder.

In one example, Clone 1 (PG13/L(tPA)SH-1), which produces tPA, was infused into the udder of a seven and a half month pregnant heifer, #90 and into the

udder of a six and a half month pregnant heifer, #56. Before infusion of the clone and particles produced by it, each quarter of the udder in both heifers was flushed with 250ml of saline as described in Example 3. After milking out the saline, the right rear (RR) quarter of heifer #90 received 2.5×10^7 cells of Clone 1 plus 80 ug/ml Polybrene in 250 ml of DMEM without serum while the right front (RF) quarter received 1×10^8 cells of Clone 1 plus 80 ug/ml Polybrene in 250 ml of DMEM without serum. The left front (LF) quarter received 250 ml of DMEM as a control while the left rear (LR) quarter was the untreated control. Heifer # 56 received 2.5×10^7 cells of Clone 1 plus 80 ug/ml Polybrene in 250 ml of DMEM without serum in the left front quarter (LF+) and 1×10^8 cells of Clone 1 plus 80 ug/ml Polybrene in 250 ml of DMEM without serum in the right rear quarter (RR+). The right front quarter received 250 ml of DMEM as a control (RF-) while the left front (LF-) quarter was the untreated control. In both heifers, the infusate was not removed and the udder was massaged three times daily to help distribute the cells and particles up into the ductwork.

15

Four weeks post-infusion, the four quarters of both heifers were stripped to remove 5 to 9ml of the viscous, serum-like premilk fluid. The premilk samples from treated and control quarters of both heifers were analyzed for the presence of tPA by Western blotting. A Western blot showing the tPA results obtained from the best quarter of each heifer is shown in Figure 2.

20

For Western blotting, the premilk samples were adjusted to pH 4.5 with acetic acid and centrifuged at $13,000 \times g$ to pellet the "curd" fraction. The "whey" fraction contained in the supernatant was collected and diluted 1:5 with sample buffer. Samples of 20 ul were loaded onto a 7.5% SDS PAGE gel. To detect tPA, the gel was transferred to nitrocellulose and then blocked overnight in 5% bovine serum albumin (BSA). The blot was incubated for 2 hours with a polyclonal antibody to tPA, #385R, purchased from American Diagnostica, diluted 1:500 with PBS followed by extensive washing in PBS. The secondary antibody, horseradish peroxidase goat-anti-rabbit, diluted 1:5000 in PBS, was incubated with the blot for one hour then extensively washed. Detection of antibody staining of tPA was by enhanced chemi-

25

30

luminescence (ECL) Amersham Detect Kit. Specificity of the primary antibody was previously determined by Western blotting a control sample of commercial tPA purchased from American Diagnostica.

5 Results of Western blotting samples obtained from the untreated control quarter (LF-) and from one treated quarters of each heifer are shown in Figure 2. Lane 1 shows a strong tPA band in the premilk from the treated (LH+) quarter of heifer #56. The middle lane, lane 2, contains premilk from the (LF-) control quarter of cow #56. Lane 3 shows a tPA band in the premilk from the treated (RF+) quarter
10 of heifer #90. A faint smear in all 3 sample lanes is the result of non-specific binding of the antibody to an unidentified protein and should be disregarded.

15 **Example 5: Determination of viability and attachment of 3T3 cells to the surface of the bovine udder.**

 The cells used in this experiment were 3T3 cells (purchased from ATCC, catalog no. CCL-92) which had been transfected with luciferase as a marker protein. These 3T3 cells are essentially identical to the PG13 packaging cells except that the 3T3 cells have not been transfected with the retroviral genes that are found in the
20 packaging line.

 Three quarters of the udder of a 5 month pregnant heifer and three quarters of the udder of a 7 month pregnant heifer were flushed with saline solution at 37° C using a standard veterinary infusion apparatus. The saline was then milked out and
25 measured to determine what quantity of medium with or without cells could be infused into that particular quarter. In both heifers, one udder quarter was left untreated as a negative control; a second quarter received only DMEM, the third quarter received 1×10^6 cells suspended in an appropriate amount of DMEM, and the fourth quarter received 1×10^7 cells also suspended in an appropriate amount of
30 DMEM. The quarters were then milked to produce about 10 ml of fluid at 3 hours, 24 hours, and 72 hours post-infusion.

The fluid obtained by milking was divided in half. One portion was cultured under standard conditions for 3T3 cells to determine if the cells were viable. The second portion was used in a luciferase assay to count the approximate number of 3T3-luciferase cells per ml in the fluid. A control was run simultaneously using known numbers of 3T3-luciferase cells and a curve drawn to determine the correlation of luciferase intensity with cell number.

The results of two experiments indicated that the highest number of cells were found 3 hours post infusion and progressively fewer cells were found in the fluid milked from the udder at 24 and 72 hours. However the cells which were milked out at 72 hours were viable. This is evidence that the cells are attaching to the udder surface and that those cells which were unattached (and therefore could be milked out of the udder) remained viable for the period of the experiment. This suggests that the 3T3 cells are not quickly destroyed by the factors in the environment of the udder. Thus PG13(tPA) cells can be expected to remain viable for at least a period of 3 days during which time they will continue to produce viral-derived particles and can therefore trans-infect a substantial number of mammary epithelial cells.

Table 1: Attachment of 3T3 cells to udder walls

| Udder Quarter | Treatment | Attachment of 3T3 cells to udder interior at: | | |
|---------------|---------------------------|---|------------------------|------------------------|
| | | 3 hours post-infusion | 24 hours post-infusion | 72 hours post-infusion |
| left hind | no treatment | - | - | - |
| left front | medium only | - | - | - |
| right hind | 1 X 10 ⁶ cells | - | +/- | + |
| right front | 1 x 10 ⁷ cells | +/- | ++ | ++ |

Example 6: Infusion of 3T3 cells grown on a solid support means

The cells used in this experiment were 3T3 cells, as described in Example 2.

In this example however the cells were grown on Cytodex beads.

The cells growing on beads were infused into the udder of a 7 month pregnant heifer and subsequently removed 1, 2 and 3 days later as described in Example 2. Cells
5 were removed from beads and viability was determined by Trypan blue exclusion. The cells remained viable for the entire period.

Example 7: Use of Silicone to displace the cells and viral-derived particles

The cells used in this experiment were 3T3 cells, as described in Example 3.
10 The volume of the solution containing the cells and viral-derived particles was reduced by 50 ml, 25 ml, or 10 ml but the overall number of the cells remained constant.

After the solution was infused into the udder, as described in the preceding
15 examples, approximately 50 ml, 25 ml or 10 ml of silicone was infused into the mammary gland, using the methods previously described.

After three days the silicone was removed from the cistern by milking. No
reaction to the silicone was noted.

20

All scientific publications and patent documents are incorporated herein by reference.

References:

25

Archer, J.S., W.S. Kennan, M.N. Gould, R.D. Bremel. 1994. Human growth hormone (hGH) secretion in milk of goats after direct transfer of the hGH gene into the mammary gland by using replication-defective retrovirus vectors. Proc. Natl. Acad. Sci. USA, 91:6840-6844.

30

Cepko, C. 1992. Transduction of genes using retrovirus vectors. *In*: Short Protocols

in Molecular Biology, 2nd edition. Ed. F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Struhl. Pub. by John Wiley & Sons. New York, N.Y. pp 9-30 to 9-45.

- 5 Furth, P.A., A. Shamay, R.J. Wall, L. Henninghausen. 1992. Gene transfer into somatic tissues by jet injection. *Analyt. Biochem.* 205:365-368.

Miller, A.D., D.G. Miller, J. V. Garcia, C.M. Lynch. 1990. Use of retroviral vectors for gene transfer and expression. *Meth. Enzymol.* 217: 581-599.

10

Morsy, M.A. and C.T. Caskey. 1997. Safe gene vectors made simpler. *Nature Biotech.* 15:17.

- 15 Pennica, D, W.E. Holmes, W.J. Kohr, R.N. Harkins, G.A. Veach, C.A. Ward, W.F. Bennett, E. Yelverton, P.H. Seeburg, H.L. Heyneker, D.V. Goeddel. 1983. Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli*. *Nature* 301: 214-221.

- 20 Thierry, A.R., Y. Lunardi-Iskandar, J.L. Bryant, P. Rabinovich, R.C. Gallo, L.C. Mahan. 1995. Systemic gene therapy: Biodistribution and long-term expression of a transgene in mice. *Proc. Natl. Acad. Sci. USA.* 92:9742-9746.

25 The present invention has been described with regard to preferred embodiments. However, it will be obvious to persons skilled in the art that a number of variations and modifications can be made without departing from the scope of the invention as described in the following claims.

30

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Her Majesty in Right of Canada as Rep. by
Agriculture and Agri-Food Canada
- (B) STREET: Experimental Farm
- (C) CITY: Ottawa
- (D) STATE: Ontario
- (E) COUNTRY: Canada
- (F) POSTAL CODE (ZIP): K1A 0C6

- (ii) TITLE OF INVENTION: Production of Value-added Milk by
Incorporation of Specific DNA Sequences into Mammary
Epithelial Cells

- (iii) NUMBER OF SEQUENCES: 4

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:

APPLICATION NUMBER: CA 0,000,000

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: CA 2,199,212
- (B) FILING DATE: 05-MAR-1997

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7699 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTGCTAG CAATTGCTAG CAATTGCTAG CAATTCATAC CAGATCACCG AAAACTGTCC 60
TCCAAATGTG TCCCCCTCAC ACTCCCAAAT TCGCGGGCTT CTGCCTCTTA GACCACTCTA 120

| | |
|---|------|
| CCCTATTCCC CACACTCACC GGAGCCAAAG CCGCGGCCCT TCCGTTTCTT TGCTTTTGAA | 180 |
| AGACCCCACC CGTAGGTGGC AAGCTAGCTT AAGTAACGCC ACTTTGCAAG GCATGGAAAA | 240 |
| ATACATAACT GAGAATAGAA AAGTTCAGAT CAAGGTCAGG AACAAAGAAA CAGCTGAATA | 300 |
| CCAAACAGGA TATCTGTGGT AAGCGGTTC TCCCCGGCT CAGGGCCAAG AACAGATGAG | 360 |
| ACAGCTGAGT GATGGGCCAA ACAGGATATC TGTGGTAAGC AGTTCCTGCC CCGGCTCGGG | 420 |
| GCCAAGAACA GATGGTCCCC AGATGCGGTC CAGCCCTCAG CAGTTTCTAG TGAATCATCA | 480 |
| GATGTTTCCA GGGTGCCCCA AGGACCTGAA AATGACCCTG TACCTTATTT GAACTAACCA | 540 |
| ATCAGTTCGC TTCTCGCTTC TGTTCGCGCG CTTCCGCTCT CCGAGCTCAA TAAAAGAGCC | 600 |
| CACAACCCCT CACTCGGCGC GCCAGTCTTC CGATAGACTG CGTCGCCCGG GTACCCGTAT | 660 |
| TCCCAATAAA GCCTCTTGCT GTTTGCATCC GAATCGTGGT CTCGCTGTTC CTTGGGAGGG | 720 |
| TCTCCTCTGA GTGATTGACT ACCCACGACG GGGGTCTTTC ATTTGGGGGC TCGTCCGGGA | 780 |
| TTTGAGAGCC CCTGCCCAGG GACCACCGAC CCACCACCGG GAGGTAAGCT GGCCAGCAAC | 840 |
| TTATCTGTGT CTGTCCGATT GTCTAGTGTC TATGTTTGAT GTTATGCGCC TCGCTCTGTA | 900 |
| CTAGTTAGCT AACTAGCTCT GTATCTGGCG GACCCGTGGT GGAAGTGACG AGTTCTGAAC | 960 |
| ACCCGGCCGC AACCTGGGA GACGTCCCAG GGACTTTGGG GGCCGTTTTT GTGGCCCGAC | 1020 |
| CTGAGGAAGG GAGTCGATGT GGAATCCGAC CCCGTCAGGA TATGTGGTTC TGGTAGGAGA | 1080 |
| CGAGAACCTA AAACAGTTCC CGCCTCCGTC TGAATTTTTG CTTTCGGTTT GGAACCGAAG | 1140 |
| CCGCGCGTCT TGTCTGCTGC AGCGCTGCAG CATCGTTCTG TGTGTCTCT GTCTGACTGT | 1200 |
| GTTTCTGTAT TTGTCTGAAA ATTAGGGCCA GACTGTTACC ACTCCCTTAA GTTTGACCTT | 1260 |
| AGGTCACTGG AAAGATGTCG AGCGGATCGC TCACAACCAG TCGGTAGATG TCAAGAAGAG | 1320 |
| ACGTTGGGTT ACCTTCTGCT CTGCAGAATG GCCAACCTTT AACGTCGGAT GGCCGCGAGA | 1380 |
| CGGCACCTTT AACCGAGACC TCATCACCCA GGTTAAGATC AAGGTCTTTT CACCTGGCCC | 1440 |
| GCATGGACAC CCAGACCAGG TCCCCTACAT CGTGACCTGG GAAGCCTTGG CTTTTGACCC | 1500 |
| CCCTCCCTGG GTCAAGCCCT TTGTACACCC TAAGCCTCCG CCTCCTCTTC CTCCATCCGC | 1560 |
| CCCGTCTCTC CCCCTTGAAC CTCCTCGTTC GACCCCGCCT CGATCCTCCC TTTATCCAGC | 1620 |
| CCTCACTCCT TCTCTAGGCG CCGGAATTCG TTAAGTCGAC ATGGAAGTCC TTCTCCTCCT | 1680 |
| CGCAGCTGTC GGGCTTTGCT GGGCACAGTA CAATCCCAAC ACTCAGGCTG GGAGGACATC | 1740 |
| TATCGTGCAT CTCTTTGAAT GGCGCTGGGC CGACATTGCA CTGGAGTGCG AACACTATTT | 1800 |

| | |
|--|------|
| AGCTCCTAAT GGGTTTGGAG GAGTTCAGGT TTCTCCTCCA AATGAAAACA TTGTCATTAC | 1860 |
| TAATCCGAAC AGGCCCTGGT GGGAAAGATA CCAGCCCATC AGCTACAAGA TCTGCAGTCG | 1920 |
| ATCGGGCAAT GAAAATGAAT TCAGAGACAT GGTGACCAGA TGCAACAATG TTGGAGTTCCG | 1980 |
| TATTTATGTG GATGCTGTTG TCAATCACAT GTGTGGATCT ATGGGTGGCA CGGGCACCCA | 2040 |
| CTCAACATGT GGGAGCTATT TCAACACCGG GACTAGAGAT TTTCCCGCTG TGCCGTACTC | 2100 |
| TGCCTGGGAT TTCAATGACG GCAAATGTCA CACTGCAAGT GGAGACATCG AAAATTATGG | 2160 |
| GGACATGTAT CAGGTCCGGG ATTGCAAGTT GTCCAGCCTT CTTGATCTGG CTCTGGAGAA | 2220 |
| GGACTATGTA CGCTCAACAA TTGCAGCGTA CATGAATCAC CTCATTGATA TGGGTGTAGC | 2280 |
| AGGGTTCCGG ATCGATGCTG CCAAGCATAT GTGGCCAGGG GACATAAGAG CGTTTCTGGA | 2340 |
| CAAACTGCAC GATCTAAATA CTCAGTGGTT TTCAGCAGGA ACGAAACCCT TTATTTACCA | 2400 |
| AGAGGTAATT GACTTGGGAG GAGAGCCAAT CACAGGCAGT CAGTACTTTG GGAATGGCCG | 2460 |
| CGTGACAGAA TTCAAGTATG GTGCCAACT GGGGACGGTG ATCCGGAAGT GGAATGGAGA | 2520 |
| GAAGATGGCC TACTTAAAGA ACTGGGGAGA AGGCTGGGGC TTTGTGCCTT CTGACAGAGC | 2580 |
| CCTGGTGTTC GTGGATAACC ACGACAACCA GCGGGGGCAC GGGGCAGGCG GAGCTTCCAT | 2640 |
| TCTTACTTTC TGGGATGCCA GGCTTTATAA AATGGCGGTT GGTTCATGC TCGCTCATCC | 2700 |
| GTACGGGTTC ACACGGGTGA TGTCAAGTTA TCGTTGGCCA AGATATTTTCG AAAACGGAGT | 2760 |
| GGATGTTAAC GACTGGGTGG GACCACCAAG TAACTCGGAC GGATCGACGA AGTCCGTTAC | 2820 |
| AATCAACGCA GACACTACCT GTGGCAATGA CTGGGTCTGC GAACATCGCT GGCGACAAAT | 2880 |
| AAGGAACATG GTTATCTTCC GTAATGTGGT AGACGGTCAG CCTTCTCAA ACTGGTGGGA | 2940 |
| CAACGGGAGC AATCAAGTAG CTTTCGGTCG CGGCGACAGA GGCTTCATTG TCTTTAATAA | 3000 |
| TGATGACTGG TATATGAATG TCGATTGCA AACTGGTCTG CCTGCTGGAA CCTACTGCGA | 3060 |
| TGTTATTCTT GGACAAAAGG AAGGCAGTGC GTGTACTGGA AAGCAGGTGT ACGTTTCCTC | 3120 |
| GGATGGAAAG GCCAATTTCC AGATTAGTAA CAGCGATGAA GATCCATTG TTGCAATTCA | 3180 |
| CGTTGATGCC AAGTTATAAG CTTTCGAGGAT CCACTAGTAA CGGCCGCCAG TGTGCTGGAA | 3240 |
| TTCGGCTTGT CGACATCTAG GCGGCCAAT TCCGCCCTC TCCCCCCCCC CCCTAACGTT | 3300 |
| ACTGGCCGAA GCCGCTTGGA ATAAGGCCGG TGTGTGTTG TCTATATGTG ATTTTCCACC | 3360 |
| ATATTGCCGT CTTTGGCAA TGTGAGGGCC CGGAAACCTG GCCCTGTCTT CTTGACGAGC | 3420 |
| ATTCCTAGGG GTCTTTCCCC TCTCGCCAAA GGAATGCAAG GTCTGTTGAA TGTCGTGAAG | 3480 |

| | |
|--|------|
| GAAGCAGTTC CTCTGGAAGC TTCTTGAAGA CAAACAACGT CTGTAGCGAC CCTTTGCAGG | 3540 |
| CAGCGGAACC CCCCACCTGG CGACAGGTGC CTCTGCGGCC AAAAGCCACG TGTATAAGAT | 3600 |
| ACACCTGCAA AGGCGGCACA ACCCCAGTGC CACGTTGTGA GTTGATAGT TGTGGAAAGA | 3660 |
| GTCAAATGGC TCTCCTCAAG CGTAGTCAAC AAGGGGCTGA AGGATGCCCA GAAGGTACCC | 3720 |
| CATTGTATGG GAATCTGATC TGGGGCCTCG GTGCACATGC TTTACATGTG TTTAGTCGAG | 3780 |
| GTTAAAAAAG CTCTAGGCCC CCCGAACCAC GGGGACGTGG TTTTCCTTTG AAAAACACGA | 3840 |
| TGATAAGCTT GCCACAACCC GGGATAATTC CTGCAGCCAA TATGGGATCG GCCATTGAAC | 3900 |
| AAGATGGATT GCACGCAGGT TCTCCGGCCG CTTGGGTGGA GAGGCTATTC GGCTATGACT | 3960 |
| GGGCACAACA GACAATCGGC TGCTCTGATG CCGCCGTGTT CCGGCTGTCA GCGCAGGGGC | 4020 |
| GCCCCGTTCT TTTGTCAAG ACCGACCTGT CCGGTGCCCT GAATGAACTG CAGGACGAGG | 4080 |
| CAGCGCGGCT ATCGTGGCTG GCCACGACGG GCGTTCCTTG CGCAGCTGTG CTCGACGTTG | 4140 |
| TCACTGAAGC GGAAGGGAC TGGCTGCTAT TGGGCGAAGT GCCGGGGCAG GATCTCCTGT | 4200 |
| CATCTCACCT TGCTCCTGCC GAGAAAGTAT CCATCATGGC TGATGCAATG CGGCGGCTGC | 4260 |
| ATACGCTTGA TCCGGCTACC TGCCCATTCTG ACCACCAAGC GAAACATCGC ATCGAGCGAG | 4320 |
| CACGTACTCG GATGGAAGCC GGTCTTGTCG ATCAGGATGA TCTGGACGAA GAGCATCAGG | 4380 |
| GGCTCGCGCC AGCCGAACTG TTCGCCAGGC TCAAGGCGCG CATGCCCGAC GGCGATGATC | 4440 |
| TCGTCGTGAC CCATGGCGAT GCCTGCTTGC CGAATATCAT GGTGGAAAAT GGCCGCTTTT | 4500 |
| CTGGATTTCAT CGACTGTGGC CGGCTGGGTG TGGCGGACCG CTATCAGGAC ATAGCGTTGG | 4560 |
| CTACCCGTGA TATTGCTGAA GAGCTTGGCG GCGAATGGGC TGACCGCTTC CTCGTGCTTT | 4620 |
| ACGGTATCGC CGCTCCCGAT TCGCAGCGCA TCGCCTTCTA TCGCCTTCTT GACGAGTTCT | 4680 |
| TCTGAGCGGG ACTCTGGGGT TCGATAAAAT AAAAGATTTT ATTTAGTCTC CAGAAAAAGG | 4740 |
| GGGGAATGAA AGACCCACCC TGTAGGTTTG GCAAGCTAGC TTAAGTAACG CCATTTTGCA | 4800 |
| AGGCATGGAA AAATACATAA CTGAGAATAG AGAAGTTCAG ATCAAGGTCA GGAACAGATG | 4860 |
| GAACAGCTGA ATATGGGCCA AACAGGATAT CTGTGGTAAG CAGTTCCTGC CCCGGCTCAG | 4920 |
| GGCCAAGAAC AGATGGAACA GCTGAATATG GGCCAAACAG GATATCTGTG GTAAGCAGTT | 4980 |
| CCTGCCCCGG CTCAGGGCCA AGAACAGATG GTCCCCAGAT GCGGTCCAGC CCTCAGCAGT | 5040 |
| TTCTAGAGAA CCATCAGATG TTTCCAGGGT GCCCAAGGA CCTGAAATGA CCCTGTGCCT | 5100 |
| TATTTGAACT AACCAATCAG TTCGCTTCTC GCTTCTGTTC GCGCGCTTCT GCTCCCCGAG | 5160 |

| | |
|--|------|
| CTCAATAAAA GAGCCCACAA CCCCTCACTC GGGGCGCCAG TCCTCCGATT GACTGAGTCG | 5220 |
| CCCCGGTACC CGTGTATCCA ATAAACCCTC TTGCAGTTGC ATCCGACTTG TGGTCTCGCT | 5280 |
| GTTCCTTGGG AGGGTCTCCT CTGAGTGATT GACTACCCGT CAGCGGGGGT CTTTCATTTG | 5340 |
| GGGGCTCGTC CGGGATCGGG AGACCCCTGC CCAGGGACCA CCGACCCACC ACCGGGAGGT | 5400 |
| AAGCTGGCTG CCTCGCGCGT TTCGGTGATG ACGGTGAAAA CCTCTGACAC ATGCAGCTCC | 5460 |
| CGGAGACGGT CACAGCTTGT CTGTAAGCGG ATGCCGGGAG CAGACAAGCC CGTCAGGGCG | 5520 |
| CGTCAGCGGG TGTGGCGGG TGTCGGGGCG CAGCCATGAC CCAGTCACGT AGCGATAGCG | 5580 |
| GAGTGATATAC TGGCTTAACT ATGCGGCATC AGAGCAGATT GTACTGAGAG TGCACCATAT | 5640 |
| GCGGTGTGAA ATACCGCACA GATGCGTAAG GAGAAAATAC CGCATCAGGC GCTCTTCCGC | 5700 |
| TTCCTCGCTC ACTGACTCGC TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA | 5760 |
| CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG | 5820 |
| AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA | 5880 |
| TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA | 5940 |
| CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCCTCTCC | 6000 |
| TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC | 6060 |
| GCTTTCTCAT AGCTCAGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT | 6120 |
| GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG | 6180 |
| TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG | 6240 |
| GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA | 6300 |
| CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG | 6360 |
| AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT | 6420 |
| TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT | 6480 |
| TTCTACGGGG TCTGACGCTC AGTGGAACGA AACTCACGT TAAGGGATTT TGGTCATGAG | 6540 |
| ATTATCAAAA AGGATCTTCA CCTAGATCCT TTAAATTAA AAATGAAGTT TTAAATCAAT | 6600 |
| CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC | 6660 |
| TATCTCAGCG ATCTGTCTAT TTCGTTTCATC CATAGTTGCC TGAATCCCCG TCGTGTAGAT | 6720 |
| AACTACGATA CGGGAGGGCT TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC | 6780 |
| ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG | 6840 |

| | |
|--|------|
| AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAAAGCTAG | 6900 |
| AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTG CAGGCATCGT | 6960 |
| GGTGTACGCG TCGTCGTTTG GTATGGCTTC ATTCAGCTCC GGTTCCTAAC GATCAAGGCG | 7020 |
| AGTTACATGA TCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT | 7080 |
| TGTCAGAAGT AAGTTGGCCG CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC | 7140 |
| TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC | 7200 |
| ATTCTGAGAA TAGTGTATGC GCGGACCGAG TTGCTCTTGC CCGGCGTCAA CACGGGATAA | 7260 |
| TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG | 7320 |
| AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC | 7380 |
| CAACTGATCT TCAGCATCTT TTAATTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG | 7440 |
| GCAAAATGCC GCAAAAAAGG GAATAAGGCG GACACGGAAA TGTGAATAC TCATACTCTT | 7500 |
| CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT | 7560 |
| TGAATGTATT TAGAAAAATA AACAAATAGG GGTCCGCGC ACATTTCCCC GAAAAGTGCC | 7620 |
| ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GCGGTATCAC | 7680 |
| GAGGCCCTTT CGTCTTCAA | 7699 |

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7980 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

| | |
|---|-----|
| GAATTGCTAG CAATTGCTAG CAATTGCTAG CAATTCATAC CAGATCACCG AAAACTGTCC | 60 |
| TCCAAATGTG TCCCCCTCAC ACTCCCAAAT TCGCGGGCTT CTGCCTCTTA GACCACTCTA | 120 |
| CCCTATTCCC CACACTCACC GGAGCCAAAG CCGCGGCCCT TCGTTTCTT TGCTTTTGAA | 180 |
| AGACCCACCC CGTAGGTGGC AAGCTAGCTT AAGTAACGCC ACTTTGCAAG GCATGGAAAA | 240 |
| ATACATAACT GAGAATAGAA AAGTTCAGAT CAAGGTCAGG AACAAAGAAA CAGCTGAATA | 300 |
| CCAAACAGGA TATCTGTGGT AAGCGGTTCC TGCCCCGGCT CAGGGCCAAG AACAGATGAG | 360 |

| | |
|--|------|
| ACAGCTGAGT GATGGGCCAA ACAGGATATC TGTGGTAAGC AGTTCCTGCC CCGGCTCGGG | 420 |
| GCCAAGAACA GATGGTCCCC AGATGCGGTC CAGCCCTCAG CAGTTTCTAG TGAATCATCA | 480 |
| GATGTTTCCA GGGTGCCCCA AGGACCTGAA AATGACCCTG TACCTTATTT GAACTAACCA | 540 |
| ATCAGTTCGC TTCTCGCTTC TGTTGCGCGC CTTCCGCTCT CCGAGCTCAA TAAAAGAGCC | 600 |
| CACAACCCCT CACTCGGCGC GCCAGTCTTC CGATAGACTG CGTCGCCCCG GTACCCGTAT | 660 |
| TCCCAATAAA GCCTCTTGCT GTTTGCATCC GAATCGTGGT CTCGCTGTTT CTTGGGAGGG | 720 |
| TCTCCTCTGA GTGATTGACT ACCCAGCAGC GGGGTCTTTC ATTTGGGGGC TCGTCCGGGA | 780 |
| TTTGAGAGACC CCTGCCCAGG GACCACCGAC CCACCACCGG GAGGTAAGCT GGCCAGCAAC | 840 |
| TTATCTGTGT CTGTCCGATT GTCTAGTGTC TATGTTTGAT GTTATGCGCC TCGCTCTGTA | 900 |
| CTAGTTAGCT AACTAGCTCT GTATCTGGCG GACCCGTGGT GGAAC TGACG AGTTCTGAAC | 960 |
| ACCCGCGCGC AACCCTGGGA GACGTCCAG GGACTTTGGG GGCCGTTTTT GTGGCCCGAC | 1020 |
| CTGAGGAAGG GAGTCGATGT GGAATCCGAC CCCGTCAGGA TATGTGGTTC TGGTAGGAGA | 1080 |
| CGAGAACCTA AAACAGTTCC CGCCTCCGTC TGAATTTTTG CTTTCGGTTT GGAACCGAAG | 1140 |
| CCGCGCGTCT TGTCTGCTGC AGCGCTGCAG CATCGTTCTG TGTTGTCTCT GTCTGACTGT | 1200 |
| GTTTCTGTAT TTGTCTGAAA ATTAGGGCCA GACTGTTACC ACTCCCTTAA GTTTGACCTT | 1260 |
| AGGTCACTGG AAAGATGTCG AGCGGATCGC TCACAACCAG TCGGTAGATG TCAAGAAGAG | 1320 |
| ACGTTGGGTT ACCTTCTGCT CTGCAGAATG GCCAACCTTT AACGTCGGAT GGCCGCGAGA | 1380 |
| CGGCACCTTT AACCGAGACC TCATCACCCA GGTAAAGATC AAGGTCTTTT CACCTGGCCC | 1440 |
| GCATGGACAC CCAGACCAGG TCCCCTACAT CGTGACCTGG GAAGCCTTGG CTTTTGACCC | 1500 |
| CCCTCCCTGG GTCAAGCCCT TTGTACACCC TAAGCCTCCG CCTCCTCTTC CTCCATCCGC | 1560 |
| CCCGTCTCTC CCCCTTGAAC CTCCTCGTTC GACCCCGCCT CGATCCTCCC TTTATCCAGC | 1620 |
| CCTCACTCCT TCTCTAGGCG CCGGAATTCG TTAACCTGAC ATGGAAGTCC TTCTCCTCCT | 1680 |
| CTCAGCTGTC GGGCTTTGCT GGGCACAGTA CAATCCCAAC ACTCAGGCTG GGAGGACATC | 1740 |
| TATCGTG CAT CTCTTTGAAT GGCGCTGGGC CGACATTGCA CTGGAGTGCG AACACTATTT | 1800 |
| AGCTCCTAAT GGGTTTGGAG GAGTTCAGGT TTCTCCTCCA AATGAAAACA TTGTCATTAC | 1860 |
| TAATCCGAAC AGGCCCTGGT GGGAAAGATA CCAGCCCATC AGCTACAAGA TCTGCAGTCG | 1920 |
| ATCGGGCAAT GAAAATGAAT TCAGAGACAT GGTGACCAGA TGCAACAATG TTGGAGTTCG | 1980 |
| TATTTATGTG GATGCTGTTG TCAATCACAT GTGTGGATCT ATGGGTGGCA CGGGCACCCA | 2040 |

| | |
|--|------|
| CTCAACATGT GGGAGCTATT TCAACACCGG GACTAGAGAT TTTCCCGCTG TGCCGTACTC | 2100 |
| TGCCTGGGAT TTCAATGACG GCAAATGTCA CACTGCAAGT GGAGACATCG AAAATTATGG | 2160 |
| GGACATGTAT CAGGTCCGGG ATTGCAAGTT GTCCAGCCTT CTTGATCTGG CTCTGGAGAA | 2220 |
| GGACTATGTA CGCTCAACAA TTGCAGCGTA CATGAATCAC CTCATTGATA TGGGTGTAGC | 2280 |
| AGGGTTCCGG ATCGATGCTG CCAAGCATAT GTGGCCAGGG GACATAAGAG CGTTTCTGGA | 2340 |
| CAAACTGCAC GATCTAAATA CTCAGTGGTT TTCAGCAGGA ACGAAACCCCT TTATTTACCA | 2400 |
| AGAGGTAATT GACTTGGGAG GAGAGCCAAT CACAGGCAGT CAGTACTTTG GGAATGGCCG | 2460 |
| CGTGACAGAA TTCAAGTATG GTGCCAACT GGGGACGGTG ATCCGGAAGT GGAATGGAGA | 2520 |
| GAAGATGGCC TACTTAAAGA ACTGGGGAGA AGGCTGGGGC TTTGTGCCTT CTGACAGAGC | 2580 |
| CCTGGTGTTT GTGGATAACC ACGACAACCA GCGGGGGCAC GGGGCAGGCG GAGCTTCCAT | 2640 |
| TCTTACTTTC TGGGATGCCA GGCTTTATAA AATGGCGGTT GGTTCATGC TCGCTCATCC | 2700 |
| GTACGGGTTC ACACGGGTGA TGTCAAGTTA TCGTTGGCCA AGATATTTG AAAACGGAGT | 2760 |
| GGATGTTAAC GACTGGGTGG GACCACCAAG TAACTCGGAC GGATCGACGA AGTCCGTTAC | 2820 |
| AATCAACGCA GACACTACCT GTGGCAATGA CTGGGTCTGC GAACATCGCT GGCACAAAT | 2880 |
| AAGGAACATG GTTATCTTCC GTAATGTGGT AGACGGTCAG CCTTCTCAA ACTGGTGGGA | 2940 |
| CAACGGGAGC AATCAAGTAG CTTTCGGTCG CGGCGACAGA GGCTTCATTG TCTTTAATAA | 3000 |
| TGATGACTGG TATATGAATG TCGATTGCA AACTGGTCTG CCTGCTGGAA CCTACTGCGA | 3060 |
| TGTTATTCTT GGACAAAAGG AAGGCAGTGC GTGTACTGGA AAGCAGGTGT ACGTTTCCTC | 3120 |
| GGATGGAAAG GCCAATTTCC AGATTAGTAA CAGCGATGAA GATCCATTTG TTGCAATTCA | 3180 |
| CGTTGATGCC AAGTTATAAG CTTGAGGAT CCACTAGTAA CGGCCGCCAG TGTGCTGGAA | 3240 |
| TTGGGCTTGT CGACATCTAG GCGGCCAAT TCCGCCCCCTC TCCCCCCCCC CCCTAACGTT | 3300 |
| ACTGGCCGAA GCCGCTTGGA ATAAGGCCGG TGTGTGTTTG TCTATATGTG ATTTTCCACC | 3360 |
| ATATTGCCGT CTTTGGCAA TGTGAGGGCC CGGAAACCTG GCCCTGTCTT CTTGACGAGC | 3420 |
| ATTCCTAGGG GTCTTTCCCC TCTCGCCAAA GGAATGCAAG GTCTGTTGAA TGTCGTGAAG | 3480 |
| GAAGCAGTTC CTCTGGAAGC TTCTGAAGA CAAACAACGT CTGTAGCGAC CCTTTGCAGG | 3540 |
| CAGCGGAACC CCCACCTGG CGACAGGTGC CTCTGCGGCC AAAAGCCACG TGTATAAGAT | 3600 |
| ACACCTGCAA AGGCGGCACA ACCCCAGTGC CACGTTGTGA GTTGGATAGT TGTGGAAAGA | 3660 |
| GTCAAATGGC TCTCCTCAAG CGTAGTCAAC AAGGGGCTGA AGGATGCCCA GAAGGTACCC | 3720 |

| | |
|---|------|
| CATTGTATGG GAATCTGATC TGGGGCCTCG GTGCACATGC TTTACATGTG TTTAGTCGAG | 3780 |
| GTAAAAAAG CTCTAGGCCC CCCGAACCAC GGGGACGTGG TTTTCCTTTG AAAAACACGA | 3840 |
| TGATAAGCTT GCCACAACCC AACAGCGTC AACAGCGTGC CGCAGATCCC GGGCAATGAG | 3900 |
| ATATGAAAA GCCTGAACTC ACCGCGACGT CTGTCGAGAA GTTTCTGATC GAAAAGTTCTG | 3960 |
| ACAGCGTCTC CGACCTGATG CAGCTCTCGG AGGGCGAAGA ATCTCGTGCT TTCAGCTTCG | 4020 |
| ATGTAGGAGG GCGTGGATAT GTCCTGCGGG TAAATAGCTG CGCCGATGGT TTCTACAAAG | 4080 |
| ATCGTTATGT TTATCGGCAC TTTGCATCGG CCGCGCTCCC GATTCCGGAA GTGCTTGACA | 4140 |
| TTGGGGAATT CAGCGAGAGC CTGACCTATT GCATCTCCCG CCGTGCACAG GGTGTCACGT | 4200 |
| TGCAAGACCT GCCTGAAACC GAACTGCCCC CTGTTCTGCA GCCGGTCGCG GAGGCCATGG | 4260 |
| ATGCGATCGC TGCGGCCGAT CTTAGCCAGA CGAGCGGGTT CGGCCCATTC GGACCGCAAG | 4320 |
| GAATCGGTCA ATACACTACA TGGCGTGATT TCATATGCGC GATTGCTGAT CCCCATGTGT | 4380 |
| ATCACTGGCA AACTGTGATG GACGACACCG TCAGTGCCTC CGTCGCGCAG GCTCTCGATG | 4440 |
| AGCTGATGCT TTGGGCCGAG GACTGCCCCG AAGTCCGGCA CCTCGTGCAC GCGGATTTCTG | 4500 |
| GCTCCAACAA TGTCTGACG GACAATGGCC GCATAACAGC GGTCATTGAC TGGAGCGAGG | 4560 |
| CGATGTTCTCGG GGATTCCCAA TACGAGGTCG CCAACATCTT CTTCTGGAGG CCGTGGTTGG | 4620 |
| CTTGTATGGA GCAGCAGACG CGCTACTTCG AGCGGAGGCA TCCGGAGCTT GCAGGATCGC | 4680 |
| CGCGGCTCCG GCGGTATATG CTCCGCATTG GTCTTGACCA ACTCTATCAG AGCTTGTTTG | 4740 |
| ACGGCAATTT CGATGATGCA GCTTGGGCGC AGGGTCGATG CGACGCAATC GTCCGATCCG | 4800 |
| GAGCCGGGAC TGTGCGGCGT ACACAAATCG CCCGCAGAAG CGCGGCCGTC TGGACCGATG | 4860 |
| GCTGTGTAGA AGTACTCGCC GATAGTGGA ACCGACGCCC CAGCACTCGT CCGAGGGCAA | 4920 |
| AGGAATAGGG GAGATGGGGG AGGCTAACTG AAACACGGAA GGGCCCGCGG GACTCTGGGG | 4980 |
| TTGATAAAA TAAAAGATTT TATTTAGTCT CCAGAAAAAG GGGGGAATGA AAGACCCAC | 5040 |
| CTGTAGGTTT GGCAAGCTAG CTTAAGTAAC GCCATTTTGC AAGGCATGGA AAAATACATA | 5100 |
| ACTGAGAATA GAGAAGTTCA GATCAAGGTC AGGAACAGAT GGAACAGCTG AATATGGGCC | 5160 |
| AAACAGGATA TCTGTGGTAA GCAGTTCCTG CCCCGGCTCA GGGCCAAGAA CAGATGGAAC | 5220 |
| AGCTGAATAT GGGCCAAACA GGATATCTGT GGTAAGCAGT TCCTGCCCCG GCTCAGGGCC | 5280 |
| AAGAACAGAT GGTCCCCAGA TGCGGTCCAG CCCTCAGCAG TTTCTAGAGA ACCATCAGAT | 5340 |
| GTTTCCAGGG TGCCCCAAGG ACCTGAAATG ACCCTGTGCC TTATTTGAAC TAACCAATCA | 5400 |

| | | | | | | |
|-------------|-------------|------------|------------|------------|------------|------|
| GTTTCGCTTCT | CGCTTCTGTT | CGCGCGCTTC | TGCTCCCCGA | GCTCAATAAA | AGAGCCCACA | 5460 |
| ACCCCTCACT | CGGGGCGCCA | GTCCTCCGAT | TGACTGAGTC | GCCCCGGTAC | CCGTGTATCC | 5520 |
| AATAAACCT | CTTGCACTG | CATCCGACTT | GTGGTCTCGC | TGTTCTTGG | GAGGGTCTCC | 5580 |
| TCTGAGTGAT | TGACTACCCG | TCAGCGGGG | TCTTTCATTT | GGGGGCTCGT | CCGGGATCGG | 5640 |
| GAGACCCCTG | CCCAGGGACC | ACCGACCCAC | CACCGGGAGG | TAAGCTGGCT | GCCTCGCGCG | 5700 |
| TTTCGGTGAT | GACGGTGAAA | ACCTCTGACA | CATGCAGCTC | CCGGAGACGG | TCACAGCTTG | 5760 |
| TCTGTAAGCG | GATGCCGGGA | GCAGACAAGC | CCGTCAGGGC | GCGTCAGCGG | GTGTTGGCGG | 5820 |
| GTGTCGGGGC | GCAGCCATGA | CCCAGTCACG | TAGCGATAGC | GGAGTGATA | CTGGCTTAAC | 5880 |
| TATGCGGCAT | CAGAGCAGAT | TGTACTGAGA | GTGCACCATA | TGCGGTGTGA | AATACCGCAC | 5940 |
| AGATGCGTAA | GGAGAAAATA | CCGCATCAGG | CGCTCTTCCG | CTTCCTCGCT | CACTGACTCG | 6000 |
| CTGCGCTCGG | TCGTTCCGGCT | GCGGCGAGCG | GTATCAGCTC | ACTCAAAGGC | GGTAATACGG | 6060 |
| TTATCCACAG | AATCAGGGGA | TAACGCAGGA | AAGAACATGT | GAGCAAAAGG | CCAGCAAAAG | 6120 |
| GCCAGGAACC | GTAAAAAGGC | CGCGTTGCTG | GCGTTTTTCC | ATAGGCTCCG | CCCCCTGAC | 6180 |
| GAGCATCACA | AAAATCGACG | CTCAAGTCAG | AGGTGGCGAA | ACCCGACAGG | ACTATAAAGA | 6240 |
| TACCAGGCGT | TTCCCCCTGG | AAGCTCCCTC | GTGCGCTCTC | CTGTTCCGAC | CCTGCCGCTT | 6300 |
| ACCGGATACC | TGTCCGCCTT | TCTCCCTTCG | GGAAGCGTGG | CGCTTCTCA | TAGCTCACGC | 6360 |
| TGTAGGTATC | TCAGTTCGGT | GTAGGTCGTT | CGCTCCAAGC | TGGGCTGTGT | GCACGAACCC | 6420 |
| CCCGTTCAGC | CCGACCGCTG | CGCCTTATCC | GGTAACTATC | GTCTTGAGTC | CAACCCGGTA | 6480 |
| AGACACGACT | TATCGCCACT | GGCAGCAGCC | ACTGGTAACA | GGATTAGCAG | AGCGAGGTAT | 6540 |
| GTAGGCGGTG | CTACAGAGTT | CTTGAAGTGG | TGGCCTAACT | ACGGCTACAC | TAGAAGGACA | 6600 |
| GTATTTGGTA | TCTGCGCTCT | GCTGAAGCCA | GTTACCTTCG | GAAAAAGAGT | TGGTAGCTCT | 6660 |
| TGATCCGGCA | AACAAACCAC | CGCTGGTAGC | GGTGGTTTTT | TTGTTTGCAA | GCAGCAGATT | 6720 |
| ACGCGCAGAA | AAAAAGGATC | TCAAGAAGAT | CCTTTGATCT | TTTCTACGGG | GTCTGACGCT | 6780 |
| CAGTGGAACG | AAAATCAGC | TTAAGGGATT | TTGGTCATGA | GATTATCAAA | AAGGATCTTC | 6840 |
| ACCTAGATCC | TTTTAAATTA | AAAATGAAGT | TTTAAATCAA | TCTAAAGTAT | ATATGAGTAA | 6900 |
| ACTTGGTCTG | ACAGTTACCA | ATGCTTAATC | AGTGAGGCAC | CTATCTCAGC | GATCTGTCTA | 6960 |
| TTTCGTTTCAT | CCATAGTTGC | CTGACTCCCC | GTCGTGTAGA | TAAGTACGAT | ACGGGAGGGC | 7020 |
| TTACCATCTG | GCCCCAGTGC | TGCAATGATA | CCGCGAGACC | CACGCTCACC | GGCTCCAGAT | 7080 |

| | |
|--|------|
| TTATCAGCAA TAAACCAGCC AGCCGGAAGG GCCGAGCGCA GAAGTGGTCC TGCAACTTTA | 7140 |
| TCCGCCTCCA TCCAGTCTAT TAATTGTTGC CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT | 7200 |
| AATAGTTTGC GCAACGTTGT TGCCATTGCT GCAGGCATCG TGGTGTACG CTCGTCGTTT | 7260 |
| GGTATGGCTT CATTAGCTC CGGTTCCCAA CGATCAAGGC GAGTTACATG ATCCCCATG | 7320 |
| TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT CCTCCGATCG TTGTCAGAAG TAAGTTGGCC | 7380 |
| GCAGTGTTAT CACTCATGGT TATGGCAGCA CTGCATAATT CTCTTACTGT CATGCCATCC | 7440 |
| GTAAGATGCT TTTCTGTGAC TGGTGAGTAC TCAACCAAGT CATTCTGAGA ATAGTGATG | 7500 |
| CGGCGACCGA GTTGCTCTTG CCCGGCGTCA ACACGGGATA ATACCGCGCC ACATAGCAGA | 7560 |
| ACTTTAAAAG TGCTCATCAT TGGAAAACGT TCTTCGGGGC GAAAACTCTC AAGGATCTTA | 7620 |
| CCGCTGTTGA GATCCAGTTC GATGTAACCC ACTCGTGCAC CCAACTGATC TTCAGCATCT | 7680 |
| TTTACTTTCA CCAGCGTTTC TGGGTGAGCA AAAACAGGAA GGCAAAATGC CGCAAAAAG | 7740 |
| GGAATAAGGG CGACACGGAA ATGTTGAATA CTCATACTCT TCCTTTTTCA ATATTATTGA | 7800 |
| AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT TTGAATGTAT TTAGAAAAAT | 7860 |
| AAACAAATAG GGGTTCCGCG CACATTTCCC CGAAAAGTGC CACCTGACGT CTAAGAAACC | 7920 |
| ATTATTATCA TGACATTAACT CTATAAAAAT AGGCGTATCA CGAGGCCCTT TCGTCTTCAA | 7980 |

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7311 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

| | |
|---|-----|
| GAATTCATAC CAGATCACCG AAAACTGTCC TCCAAATGTG TCCCCCTCAC ACTCCCAAAT | 60 |
| TCGCGGGCTT CTGCCTCTTA GACCACTCTA CCCTATTCCC CACACTCACC GGAGCCAAAG | 120 |
| CCGCGGCCCT TCCGTTTCTT TGCTTTTGAA AGACCCACCG CGTAGGTGGC AAGCTAGCTT | 180 |
| AAGTAACGCC ACTTTGCAAG GCATGGAAAA ATACATAACT GAGAATAGAA AAGTTCAGAT | 240 |
| CAAGGTCAGG AACAAAGAAA CAGCTGAATA CCAAACAGGA TATCTGTGGT AAGCGGTTCC | 300 |

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|------|
| TGCCCCGGCT | CAGGGCCAAG | AACAGATGAG | ACAGCTGAGT | GATGGGCCAA | ACAGGATATC | 360 |
| TGTGGTAAGC | AGTTCCTGCC | CCGGCTCGGG | GCCAAGAACA | GATGGTCCCC | AGATGCGGTC | 420 |
| CAGCCCTCAG | CAGTTTCTAG | TGAATCATCA | GATGTTTCCA | GGGTGCCCCA | AGGACCTGAA | 480 |
| AATGACCCTG | TACCTTATTT | GAAC TAACCA | ATCAGTTTCG | TTCTCGCTTC | TGTTGCGCGC | 540 |
| CTTCCGCTCT | CCGAGCTCAA | TAAAAGAGCC | CACAACCCCT | CACTCGGCGC | GCCAGTCTTC | 600 |
| CGATAGACTG | CGTCGCCCCG | GTACCCGTAT | TCCCAATAAA | GCCTCTTGCT | GTTTGCATCC | 660 |
| GAATCGTGGT | CTCGCTGTTC | CTTGGGAGGG | TCTCCTCTGA | GTGATTGACT | ACCCACGACG | 720 |
| GGGGTCTTTC | ATTTGGGGGC | TCGTCCGGGA | TTTGAGAGCC | CCTGCCCAGG | GACCACCGAC | 780 |
| CCACCACCGG | GAGGTAAGCT | GGCCAGCAAC | TTATCTGTGT | CTGTCCGATT | GTCTAGTGTC | 840 |
| TATGTTTGAT | GTTATGCGCC | TGCGTCTGTA | CTAGTTAGCT | AACTAGCTCT | GTATCTGGCG | 900 |
| GACCCGTGGT | GGAAC TGACG | AGTTCTGAAC | ACCCGGCCGC | AACCCTGGGA | GACGTCCAG | 960 |
| GGACTTTGGG | GGCCGTTTTT | GTGGCCCGAC | CTGAGGAAGG | GAGTCGATGT | GGAATCCGAC | 1020 |
| CCCGTCAGGA | TATGTGGTTC | TGGTAGGAGA | CGAGAACCTA | AAACAGTTCC | CGCCTCCGTC | 1080 |
| TGAATTTTGT | CTTTCGGTTT | GGAACCGAAG | CCGCGCGTCT | TGTCTGCTGC | AGCGCTGCAG | 1140 |
| CATCGTTCTG | TGTTGTCTCT | GTCTGACTGT | GTTTCTGTAT | TTGTCTGAAA | ATTAGGGCCA | 1200 |
| GACTGTTACC | ACTCCCTTAA | GTTTGACCTT | AGGTCACTGG | AAAGATGTCG | AGCGGATCGC | 1260 |
| TCACAACCAG | TCGGTAGATG | TCAAGAAGAG | ACGTGGGGTT | ACCTTCTGCT | CTGCAGAATG | 1320 |
| GCCAACCTTT | AACGTCGGAT | GGCCGCGAGA | CGGCACCTTT | AACCGAGACC | TCATCACCCA | 1380 |
| GGTTAAGATC | AAGGTCTTTT | CACCTGGCCC | GCATGGACAC | CCAGACCAGG | TCCCCTACAT | 1440 |
| CGTGACCTGG | GAAGCCTTGG | CTTTTGACCC | CCCTCCCTGG | GTCAAGCCCT | TTGTACACCC | 1500 |
| TAAGCCTCCG | CCTCCTCTTC | CTCCATCCGC | CCCGTCTCTC | CCCCTTGAAC | CTCCTCGTTC | 1560 |
| GACCCCGCCT | CGATCCTCCC | TTTATCCAGC | CCTCACTCCT | TCTCTAGGCG | CCGGAATTCC | 1620 |
| GATCTGATCA | AGAGACAGGA | TGAGGATCGT | TTCGCATGAT | TGAACAAGAT | GGATTGCACG | 1680 |
| CAGGTTCTCC | GGCCGCTTGG | GTGGAGAGGC | TATTCGGCTA | TGACTGGGCA | CAACAGACAA | 1740 |
| TCGGCTGCTC | TGATGCCGCC | GTGTTCCGGC | TGTCAGCGCA | GGGGCGCCCG | GTTCTTTTGT | 1800 |
| TCAAGACCGA | CCTGTCCGGT | GCCCTGAATG | AACTGCAGGA | CGAGGCAGCG | CGGCTATCGT | 1860 |
| GGCTGGCCAC | GACGGGCGTT | CCTTGCGCAG | CTGTGCTCGA | CGTTGTCACT | GAAGCGGGAA | 1920 |
| GGGACTGGCT | GCTATTGGGC | GAAGTGCCGG | GGCAGGATCT | CCTGTCACTT | CACCTTGCTC | 1980 |

| | |
|--|------|
| CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG | 2040 |
| CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG | 2100 |
| AAGCCGGTCT TGTGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG | 2160 |
| AACTGTTTCG CAGGCTCAAG GCGCGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG | 2220 |
| GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT | 2280 |
| GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG | 2340 |
| CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCTCGT GCTTTACGGT ATCGCCGCTC | 2400 |
| CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT | 2460 |
| GGGGTTCGAA ATGACCGACC AAGCGACGCC CAACCTGCCA TCACGAGATT TCGATTCCAC | 2520 |
| CGCCGCCTTC TATGAAAGGT TGGGCTTCGG AATCGTTTTTC CGGGACGCCG GCTGGATGAT | 2580 |
| CCTCCAGCGC GGGGATCTCA TGCTGGAGTT CTTCGCCCAC CCCGGGCTCG ATCCCCTCGC | 2640 |
| GAGTTGGTTC AGCTGCTGCC TGAGGCTGGA CGACCTCGCG GAGTTCTACC GGCAGTGCAA | 2700 |
| ATCCGTCGCG ATCCAGGAAA CCAGCAGCGG CTATCCGCGC ATCCATGCCC CCGAACTGCA | 2760 |
| GGAGTGGGGA GGCACGATGG CCGCTTTGGT CGAGGCGGAT CCGGGCAGAA ATGGTTGAAC | 2820 |
| TCCCAGAGAT GTCCTACACC TAGGGGAGAA GCAGCCAAGG GGTGTGTTCC CACCAAGGAC | 2880 |
| GACCCGTCTG CGCACAAACG GATGAGCCCA TCAGACAAAG ACATATTCAT TCTCTGCTGC | 2940 |
| AAACTTGGCA TAGCTCTGCT TTGCCTGGGG CTATTGGGGG AAGTTGCGGT TCGTGCTCGC | 3000 |
| AGGGCTCTCA CCCTTGACTC TTTTAATAGC TCTTCTGTGC AAGATTACAA TCTAAACAAT | 3060 |
| TCGGAGAACT CGACCTTCCT CCTGAGGCAA GGACCACAGC CAACTTCCTC TTACAAGCCG | 3120 |
| CATCGATTTT GTCCTTCAGA AATAGAAATA AGAATGCTTG CTAAAAATTA TATTTTTACC | 3180 |
| AATAAGACCA ATCCAATAGG TAGATTATTA GTTACTATGT TAAGAAATGA ATCATTATCT | 3240 |
| TTTAGTACTA TTTTACTCA AATTCAGAAG TTAGAAATGG GAATAGAAAA TAGAAAGAGA | 3300 |
| CGCTCAACCT CAATTGAAGA ACAGGTGCAA GGACTATTGA CCACAGGCCT AGAAGTAAAA | 3360 |
| AAGGGAAAAA AGAGTGTTTT TGTCAAATA GGAGACAGGT GGTGGCAACC AGGGACTTAT | 3420 |
| AGGGGACCTT ACATCTACAG ACCAACAGAT GCCCCCTTAC CATATACAGG AAGATATGAC | 3480 |
| TTAAATTGGG ATAGGTGGGT TACAGTCAAT GGCTATAAAG TGTATATAG ATCCCTCCCT | 3540 |
| TTTCGTGAAA GACTCGCCAG AGCTAGACCT CCTTGGTGTA TGTGTCTCA AGAAGAAAAA | 3600 |
| GACGACATGA AACAACAGGT ACATGATTAT ATTTATCTAG GAACAGGAAT GCACTTTTGG | 3660 |

| | |
|---|------|
| GGAAAGATTT TCCATACCAA GGAGGGGACA GTGGCTGGAC TAATAGAACA TTATTCTGCA | 3720 |
| AAAACATCATG GCATGAGTTA TTATGAATAG CCTTTATTGG CCCAACCTTG CGGTTCCAG | 3780 |
| GGCTTAAGTA AGTTTTTGGT TACAAACTGT TCTTAAACG AGGATGTGAG ACAAGTGTT | 3840 |
| TCCTGACTTG GTTTGGTATC AAAGGTTCTG ATCTGAGCTC TGAGTGTCT ATTTTCTAT | 3900 |
| GTTCTTTTGG AATTTATCCA AATCTTATGT AAATGCTTAT GTAAACCAAG ATATAAAGA | 3960 |
| GTGCTGATTT TTTGAGTAA CTGCAACAG TCCTAACATT CACCTCTGT GTGTTTGTGT | 4020 |
| CTGTTGCGCA TCCCGTCTCC GCTCGTCACT TATCCTTCAC TTTCCAGAGG GTCCCCCGC | 4080 |
| AGACCCCGGC GACCCTCAGG TCGGCCGACT GCGGCAGCTG GCGCCCGAAC AGGGACCCTC | 4140 |
| GGATAAGTGA CCCTGTCTC TATTTCTACT ATTTGGTGT TGTCTGTAT TGTCTCTTC | 4200 |
| TTGTCTGGCT ATCATCACA GAGCGGAACG GACTCACCAT AGGGACCAAG CTTGTCGACA | 4260 |
| TTTCTGCAGA TATCCATCAC ACTGGCGGCC GCTCGAGCAT GCATCTAGAA CATCGATAA | 4320 |
| ATAAAGATT TTATTAGTC TCCAGAAAA GGGGGGAATG AAAGACCCA CCTGTAGTT | 4380 |
| TGGCAAGCTA GCTTAAGTAA CGCCATTTG CAAGGCATG AAAAATACAT AACTGAGAAT | 4440 |
| AGAGAAGTTC AGATCAAGGT CAGGAACAGA TGAACAGCT GAATATGGGC CAAACAGGAT | 4500 |
| ATCTGTGGTA AGCAGTTCCT GCGCCGCTC AGGGCCAAGA ACAGATGGAA CAGCTGAATA | 4560 |
| TGGGCCAAAC AGGATATCTG TGGTAAGCAG TTCCTGCCCC GGCTCAGGGC CAAGAACAGA | 4620 |
| TGGTCCCCAG ATGCGGTCCA GCCCTCAGCA GTTCTAGAG AACCATCAGA TGTTTCCAGG | 4680 |
| GTGCCCCAAG GACCTGAAAT GACCCTGTGC CTTATTTGAA CTAACCAATC AGTTCGCTTC | 4740 |
| TCGCTTCTGT TCGCGCGCTT CTGCTCCCCG AGCTCAATAA AAGAGCCCAC AACCCTCAC | 4800 |
| TCGGGGCGCC AGTCCTCCGA TTGACTGAGT CGCCCGGTA CCCGTGTATC CAATAAACC | 4860 |
| TCTTGCAAGT GCATCCGACT TGTGGTCTCG CTGTTCCCTG GGAGGGTCTC CTCTGAGTGA | 4920 |
| TTGACTACCC GTCAGCGGG GTCTTTCATT TGGGGGCTCG TCCGGGATCG GGAGACCCT | 4980 |
| GCCCAGGGAC CACCGACCA CCACCGGGAG GTAAGCTGGC TGCCTCGCGC GTTTCGGTGA | 5040 |
| TGACGGTGAA AACCTCTGAC ACATGCAGCT CCCGGAGACG GTCACAGCTT GTCTGTAAGC | 5100 |
| GGATGCCGGG AGCAGACAAG CCCGTCAGGG CGCGTCAGCG GGTGTTGGCG GGTGTCGGG | 5160 |
| CGCAGCCATG ACCCAGTCAC GTAGCGATAG CGGAGTGTAT ACTGGCTTAA CTATGCGGCA | 5220 |
| TCAGAGCAGA TTGTACTGAG AGTGACCAT ATGCGGTGTG AAATACCGCA CAGATGCGTA | 5280 |
| AGGAGAAAAT ACCGCATCAG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG | 5340 |

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|------|
| GTCGTTTCGGC | TGCGGCGAGC | GGTATCAGCT | CACTCAAAGG | CGGTAATACG | GTTATCCACA | 5400 |
| GAATCAGGGG | ATAACGCAGG | AAAGAACATG | TGAGCAAAAG | GCCAGCAAAA | GGCCAGGAAC | 5460 |
| CGTAAAAAGG | CCGCGTTGCT | GGCGTTTTTC | CATAGGCTCC | GCCCCCTGA | CGAGCATCAC | 5520 |
| AAAAATCGAC | GCTCAAGTCA | GAGGTGGCGA | AACCCGACAG | GACTATAAAG | ATACCAGGCG | 5580 |
| TTTCCCCCTG | GAAGCTCCCT | CGTGCGCTCT | CCTGTTCCGA | CCCTGCCGCT | TACCGGATAC | 5640 |
| CTGTCCGCCT | TTCTCCCTTC | GGGAAGCGTG | GCGCTTTCTC | ATAGCTCACG | CTGTAGGTAT | 5700 |
| CTCAGTTCGG | TGTAGGTCGT | TCGCTCCAAG | CTGGGCTGTG | TGCACGAACC | CCCCGTTTCA | 5760 |
| CCCGACCGCT | GCGCCTTATC | CGGTAACTAT | CGTCTTGAGT | CCAACCCGGT | AAGACACGAC | 5820 |
| TTATCGCCAC | TGGCAGCAGC | CACTGGTAAC | AGGATTAGCA | GAGCGAGGTA | TGTAGGCGGT | 5880 |
| GCTACAGAGT | TCTTGAAGTG | GTGGCCTAAC | TACGGCTACA | CTAGAAGGAC | AGTATTTGGT | 5940 |
| ATCTGCGCTC | TGCTGAAGCC | AGTTACCTTC | GGAAAAGAG | TTGGTAGCTC | TTGATCCGGC | 6000 |
| AAACAAACCA | CCGCTGGTAG | CGGTGGTTTT | TTGTTTGCA | AGCAGCAGAT | TACGCGCAGA | 6060 |
| AAAAAAGGAT | CTCAAGAAGA | TCCTTTGATC | TTTTCTACGG | GGTCTGACGC | TCAGTGAAC | 6120 |
| GAAAACTCAC | GTTAAGGGAT | TTTGGTCATG | AGATTATCAA | AAAGGATCTT | CACCTAGATC | 6180 |
| CTTTTAAATT | AAAAATGAAG | TTTTAAATCA | ATCTAAAGTA | TATATGAGTA | AACTTGGTCT | 6240 |
| GACAGTTACC | AATGCTTAAT | CAGTGAGGCA | CCTATCTCAG | CGATCTGTCT | ATTTTCGTTCA | 6300 |
| TCCATAGTTG | CCTGACTCCC | CGTCGTGTAG | ATAACTACGA | TACGGGAGGG | CTTACCATCT | 6360 |
| GGCCCCAGTG | CTGCAATGAT | ACCGCGAGAC | CCACGCTCAC | CGGCTCCAGA | TTTATCAGCA | 6420 |
| ATAAACCAGC | CAGCCGGAAG | GGCCGAGCGC | AGAAGTGGTC | CTGCAACTTT | ATCCGCCTCC | 6480 |
| ATCCAGTCTA | TTAATTGTTG | CCGGGAAGCT | AGAGTAAGTA | GTTCGCCAGT | TAATAGTTTG | 6540 |
| CGCAACGTTG | TTGCCATTGC | TGCAGGCATC | GTGGTGTAC | GCTCGTCGTT | TGGTATGGCT | 6600 |
| TCATTCAGCT | CCGGTTCCCA | ACGATCAAGG | CGAGTTACAT | GATCCCCCAT | GTTGTGCAAA | 6660 |
| AAAGCGGTTA | GCTCCTTCGG | TCCTCCGATC | GTTGTCAGAA | GTAAGTTGGC | CGCAGTGTTA | 6720 |
| TCACTCATGG | TTATGGCAGC | ACTGCATAAT | TCTCTTACTG | TCATGCCATC | CGTAAGATGC | 6780 |
| TTTCTGTGTA | CTGGTGAGTA | CTCAACCAAG | TCATTCTGAG | AATAGTGTAT | GCGGCGACCG | 6840 |
| AGTTGCTCTT | CCCCGGCGTC | AACACGGGAT | AATACCGCGC | CACATAGCAG | AACTTTAAAA | 6900 |
| GTGCTCATCA | TTGAAAAACG | TTCTTCGGGG | CGAAAACTCT | CAAGGATCTT | ACCGCTGTTG | 6960 |
| AGATCCAGTT | CGATGTAACC | CACTCGTGCA | CCCAACTGAT | CTTCAGCATC | TTTTACTTTC | 7020 |

| | |
|---|------|
| ACCAGCGTTT CTGGGTGAGC AAAACAGGA AGGCAAAATG CCGCAAAAAA GGAATAAGG | 7080 |
| GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG AAGCATTTAT | 7140 |
| CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA | 7200 |
| GGGGTTCGCG GCACATTTCC CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC | 7260 |
| ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTTCA A | 7311 |

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7885 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

| | |
|---|-----|
| GAATTCATAC CAGATCACCG AAAACTGTCC TCCAAATGTG TCCCCCTCAC ACTCCCAAAT | 60 |
| TCGCGGGCTT CTGCCTCTTA GACCACTCTA CCCTATTCCC CACACTCACC GGAGCCAAAG | 120 |
| CCGCGGCCCT TCCGTTTCTT TGCTTTTGAA AGACCCACCC CGTAGGTGGC AAGCTAGCTT | 180 |
| AAGTAACGCC ACTTTGCAAG GCATGGAAAA ATACATAACT GAGAATAGAA AAGTTCAGAT | 240 |
| CAAGGTCAGG AACAAAGAAA CAGCTGAATA CCAAACAGGA TATCTGTGGT AAGCGTTTCC | 300 |
| TGCCCCGGCT CAGGGCCAAG AACAGATGAG ACAGCTGAGT GATGGGCCAA ACAGGATATC | 360 |
| TGTGGTAAGC AGTTCCTGCC CCGGCTCGGG GCCAAGAACA GATGGTCCCC AGATGCGGTC | 420 |
| CAGCCCTCAG CAGTTTCTAG TGAATCATCA GATGTTTCCA GGGTGCCCCA AGGACCTGAA | 480 |
| AATGACCCTG TACCTTATTT GAACTAACCA ATCAGTTCGC TTCTCGCTTC TGTTGCGCGC | 540 |
| CTTCCGCTCT CCGAGCTCAA TAAAGAGCC CACAACCCCT CACTCGGCGC GCCAGTCTTC | 600 |
| CGATAGACTG CGTCGCCCCG GTACCCGTAT TCCCAATAAA GCCTCTTGCT GTTTGCATCC | 660 |
| GAATCGTGGT CTCGCTGTTC CTTGGGAGGG TCTCCTCTGA GTGATTGACT ACCCAGCAGC | 720 |
| GGGGTCTTTC ATTTGGGGGC TCGTCCGGA TTTGGAGACC CCTGCCCAGG GACCACCGAC | 780 |
| CCACCACCGG GAGGTAAGCT GGCCAGCAAC TTATCTGTGT CTGTCCGATT GTCTAGTGTC | 840 |
| TATGTTTGAT GTTATGCGCC TCGTCTGTA CTAGTTAGCT AACTAGCTCT GTATCTGGCG | 900 |
| GACCCGTGGT GGAAGTACG AGTTCTGAAC ACCCGGCCGC AACCTGGGA GACGTCCCAG | 960 |

| | |
|---|------|
| GGACTTTGGG GGCCGTTTTT GTGGCCCGAC CTGAGGAAGG GAGTCGATGT GGAATCCGAC | 1020 |
| CCCGTCAGGA TATGTGGTTC TGGTAGGAGA CGAGAACCTA AAACAGTTCC CGCCTCCGTC | 1080 |
| TGAATTTTTG CTTTCGGTTT GGAACCGAAG CCGCGCGTCT TGTCTGCTGC AGCGCTGCAG | 1140 |
| CATCGTTCTG TGTGTCTCT GTCTGACTGT GTTTCTGTAT TTGTCTGAAA ATTAGGGCCA | 1200 |
| GACTGTTACC ACTCCCTTAA GTTTGACCTT AGGTCACTGG AAAGATGTCG AGCGGATCGC | 1260 |
| TCACAACCAG TCGGTAGATG TCAAGAAGAG ACGTTGGGTT ACCTTCTGCT CTGCAGAATG | 1320 |
| GCCAACCTTT AACGTCGGAT GGCCGCGAGA CGGCACCTTT AACCGAGACC TCATCACCCA | 1380 |
| GGTTAAGATC AAGGTCTTTT CACCTGGCCC GCATGGACAC CCAGACCAGG TCCCCTACAT | 1440 |
| CGTGACCTGG GAAGCCTTGG CTTTGGACCC CCCTCCCTGG GTCAAGCCCT TTGTACACCC | 1500 |
| TAAGCCTCCG CCTCCTCTTC CTCCATCCGC CCCGTCTCTC CCCCTGAAC CTCCTCGTTC | 1560 |
| GACCCCGCCT CGATCCTCCC TTTATCCAGC CCTCACTCCT TCTCTAGGCG CCGGAATTCC | 1620 |
| GATCTGATCA AGAGACAGGA TGAGGATCGT TTCGCATGAT TGAACAAGAT GGATTGCACG | 1680 |
| CAGGTTCTCC GGCCGCTTGG GTGGAGAGGC TATTCGGCTA TGAAGGGCA CAACAGACAA | 1740 |
| TCGGCTGCTC TGATGCCGCC GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG | 1800 |
| TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT | 1860 |
| GGCTGGCCAC GACGGGCGTT CTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA | 1920 |
| GGGACTGGCT GCTATTGGGC GAAGTGCCGG GGAGGATCT CCTGTCTCT CACCTTGCTC | 1980 |
| CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG | 2040 |
| CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG | 2100 |
| AAGCCGGTCT TGTGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG | 2160 |
| AACTGTTGCG CAGGCTCAAG GCGCGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG | 2220 |
| GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTCTGGA TTCATCGACT | 2280 |
| GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG | 2340 |
| CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC | 2400 |
| CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT | 2460 |
| GGGGTTCGAA ATGACCGACC AAGCGACGCC CAACCTGCCA TCACGAGATT TCGATTCCAC | 2520 |
| CGCCGCCCTC TATGAAAGGT TGGGCTTCGG AATCGTTTTT CGGGACGCCG GCTGGATGAT | 2580 |
| CCTCCAGCGC GGGGATCTCA TGCTGGAGTT CTTGCCCCAC CCCGGGCTCG ATCCCCTCGC | 2640 |

| | |
|---|------|
| GAGTTGGTTC AGCTGCTGCC TGAGGCTGGA CGACCTCGCG GAGTTCTACC GGCAGTGCAA | 2700 |
| ATCCGTCGGC ATCCAGGAAA CCAGCAGCGG CTATCCGCGC ATCCATGCCC CCGAACTGCA | 2760 |
| GGAGTGGGGA GGCACGATGG CCGCTTTGGT CGAGGCGGAT CCGGGCAGAA ATGTTGAAC | 2820 |
| TCCCAGAGT GTCCTACACC TAGGGGAGAA GCAGCCAAGG GGTGTGTTCC CACCAAGGAC | 2880 |
| GACCCGTCTG CGCACAAACG GATGAGCCCA TCAGACAAAG ACATATTCAT TCTCTGCTGC | 2940 |
| AAACTTGCCA TAGCTCTGCT TTGCCTGGGG CTATTGGGGG AAGTTGCGGT TCGTGCTCGC | 3000 |
| AGGGCTCTCA CCCTTGACTC TTTTAATAGC TCTTCTGTGC AAGATTACAA TCTAAACAAT | 3060 |
| TCGGAGAACT CGACCTTCCT CCTGAGGCAA GGACCACAGC CAACTTCCTC TTACAAGCCG | 3120 |
| CATCGATTTT GTCCTTCAGA AATAGAAATA AGAATGCTTG CTAAAAATTA TATTTTTACC | 3180 |
| AATAAGACCA ATCCAATAGG TAGATTATTA GTTACTATGT TAAGAAATGA ATCATTATCT | 3240 |
| TTTAGTACTA TTTTACTCA AATTCAGAAG TTAGAAATGG GAATAGAAAA TAGAAAGAGA | 3300 |
| CGCTCAACCT CAATTGAAGA ACAGGTGCAA GGACTATTGA CCACAGGCCT AGAAGTAAAA | 3360 |
| AAGGGAAAAA AGAGTGTTTT TGTCAAATA GGAGACAGGT GGTGGCAACC AGGGACTTAT | 3420 |
| AGGGGACCTT ACATCTACAG ACCAACAGAT GCCCCCTTAC CATATACAGG AAGATATGAC | 3480 |
| TTAAATTGGG ATAGGTGGGT TACAGTCAAT GGCTATAAAG TGTATATAG ATCCCTCCCT | 3540 |
| TTTCGTGAAA GACTCGCCAG AGCTAGACCT CCTTGGTGTA TGTGTCTCA AGAAGAAAAA | 3600 |
| GACGACATGA AACAACAGGT ACATGATTAT ATTTATCTAG GAACAGGAAT GCACTTTTGG | 3660 |
| GGAAAGATTT TCCATACCAA GGAGGGGACA GTGGCTGGAC TAATAGAACA TTATTCTGCA | 3720 |
| AAAATCATG GCATGAGTTA TTATGAATAG CCTTTATTGG CCCAACCTTG CGGTTCCAG | 3780 |
| GGCTTAAGTA AGTTTTTGGT TACAACTGT TCTTAAACG AGGATGTGAG ACAAGTGGTT | 3840 |
| TCCTGACTTG GTTTGGTATC AAAGGTTCTG ATCTGAGCTC TGAGTGTCTT ATTTTCCTAT | 3900 |
| GTTCTTTTGG AATTATCCA AATCTTATGT AAATGCTTAT GTAAACCAAG ATATAAAGA | 3960 |
| GTGCTGATTT TTTGAGTAA CTTGCAACAG TCCTAACATT CACCTCTTGT GTGTTTGTGT | 4020 |
| CTGTTGCCA TCCCGTCTCC GCTCGTCACT TATCCTTCAC TTTCCAGAGG GTCCCCCGC | 4080 |
| AGACCCCGGC GACCCTCAGG TCGGCCGACT GCGGCAGCTG GCGCCGAAC AGGGACCCTC | 4140 |
| GGATAAGTGA CCCTGTCTC TATTTCTACT ATTTGGTGTT TGTCTGTAT TGTCTCTTC | 4200 |
| TTGTCTGGCT ATCATCACA GAGCGGAACG GACTCACCAT AGGGACCAAG CTTGTCGACA | 4260 |
| TCTAGGGCGG CCAATTCCGC CCCTCTCCCT CCCCCCCCC TAACGTTACT GGCCGAAGCC | 4320 |

| | |
|--|------|
| GCTTGAATA AGGCCGGTGT GCGTTTGTCT ATATGTGATT TTCCACCATA TTGCCGTCTT | 4380 |
| TTGGCAATGT GAGGGCCCGG AAACCTGGCC CTGTCTTCTT GACGAGCATT CCTAGGGGTC | 4440 |
| TTTCCCCTCT CGCCAAAGGA ATGCAAGGTC TGTGAATGT CGTGAAGGAA GCAGTTCCTC | 4500 |
| TGAAGCTTC TTGAAGACAA ACAACGCTCG TAGCGACCCT TTGCAGGCAG CGGAACCCCC | 4560 |
| CACCTGGCGA CAGGTGCCTC TGCGGCCAAA AGCCACGTGT ATAAGATACA CCTGCAAAGG | 4620 |
| CGGCACAACC CCAGTGCCAC GTTGTGAGTT GGATAGTTGT GGAAAGAGTC AAATGGCTCT | 4680 |
| CCTCAAGCGT ATTCAACAAG GGGCTGAAGG ATGCCCAGAA GGTACCCCAT TGTATGGGAT | 4740 |
| CTGATCTGGG GCCTCGGTGC ACATGCTTTA CATGTGTTTA GTCGAGGTTA AAAAAACGTC | 4800 |
| TAGGCCCCC GAACCACGGG GACGTGGTTT TCCTTTGAAA AACACGCCAA TAATATGGGC | 4860 |
| GGCCGCTCGA GCATGCATCT AGAACATCGA TAAAATAAAA GATTTTATTT AGTCTCCAGA | 4920 |
| AAAAGGGGGG AATGAAAGAC CCCACCTGTA GGTTCGGCAA GCTAGCTTAA GTAACGCCAT | 4980 |
| TTTGCAAGGC ATGGAAAAAT ACATACTGA GAATAGAGAA GTTCAGATCA AGGTCAGGAA | 5040 |
| CAGATGGAAC AGCTGAATAT GGGCCAAACA GGATATCTGT GGTAAGCAGT TCCTGCCCCG | 5100 |
| GCTCAGGGCC AAGAACAGAT GGAACAGCTG AATATGGGCC AAACAGGATA TCTGTGGTAA | 5160 |
| GCAGTTCCTG CCCC GGCTCA GGGCCAAGAA CAGATGGTCC CCAGATGCGG TCCAGCCCTC | 5220 |
| AGCAGTTTCT AGAGAACCAT CAGATGTTTC CAGGGTGCCC CAAGGACCTG AAATGACCCT | 5280 |
| GTGCCTTATT TGAAC TAACC AATCAGTTCG CTTCTCGCTT CTGTTGCGGC GCTTCTGCTC | 5340 |
| CCCGAGCTCA ATAAAAGAGC CCACAACCCC TCACTCGGGG CGCCAGTCCT CCGATTGACT | 5400 |
| GAGTCGCCCC GGTACCCGTG TATCCAATAA ACCCTCTTGC AGTTGCATCC GACTTGTGGT | 5460 |
| CTCGCTGTTC CTGGGAGGG TCTCCTCTGA GTGATTGACT ACCCGTCAGC GGGGGTCTTT | 5520 |
| CATTTGGGGG CTCGTCCGGG ATCGGGAGAC CCCTGCCCAG GGACCACCGA CCCACCACCG | 5580 |
| GGAGGTAAGC TGGCTGCCTC GCGCGTTTCG GTGATGACGG TGAAAACCTC TGACACATGC | 5640 |
| AGCTCCCGGA GACGGTCACA GCTTGTCTGT AAGCGGATGC CGGGAGCAGA CAAGCCCGTC | 5700 |
| AGGGCGCGTC AGCGGGTGTG GCGGGGTGTC GGGGCGCAGC CATGACCCAG TCACGTAGCG | 5760 |
| ATAGCGGAGT GTATACTGGC TTAAC TATGC GGCATCAGAG CAGATTGTAC TGAGAGTGCA | 5820 |
| CCATATGCGG TGTGAAATAC CGCACAGATG CGTAAGGAGA AAATACCGCA TCAGGCGCTC | 5880 |
| TTCCGCTTCC TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC | 5940 |
| AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA | 6000 |

| | |
|---|------|
| CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT | 6060 |
| TTTCCATAGG CTCCGCCCCC CTGACGAGCA TCACAAAAT CGACGCTCAA GTCAGAGGTG | 6120 |
| GCGAAACCCG ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG | 6180 |
| CTCTCCTGTT CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTGCGGAAG | 6240 |
| CGTGGCGCTT TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC | 6300 |
| CAAGCTGGGC TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA | 6360 |
| CTATCGTCTT GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG | 6420 |
| TAACAGGATT AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC | 6480 |
| TAACACGGC TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC | 6540 |
| CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAAACA ACCACCGCTG GTAGCGGTGG | 6600 |
| TTTTTTTGTG TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT | 6660 |
| GATCTTTTCT ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT | 6720 |
| CATGAGATTA TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA | 6780 |
| ATCAATCTAA AGTATATATG AGTAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA | 6840 |
| GGCACCTATC TCAGCGATCT GTCTATTTG TCCATCCATA GTTGCCCTGAC TCCCCGTCGT | 6900 |
| GTAGATAACT ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG | 6960 |
| AGACCCACGC TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA | 7020 |
| GCGCAGAAGT GGTCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA | 7080 |
| AGCTAGAGTA AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTGCGCA TTGCTGCAGG | 7140 |
| CATCGTGGTG TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC | 7200 |
| AAGGCGAGTT ACATGATCCC CCATGTTGTG CAAAAAGCG GTTAGCTCCT TCGGTCCTCC | 7260 |
| GATCGTTGTC AGAAGTAAGT TGGCCGAGT GTTATCACTC ATGGTTATGG CAGCACTGCA | 7320 |
| TAATTCTCTT ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC | 7380 |
| CAAGTCATTC TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG CGTCAACACG | 7440 |
| GGATAATACC GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTTGAA AACGTTCTTC | 7500 |
| GGGGCGAAAA CTCTCAAGGA TCTTACCCT GTTGAGATCC AGTTCGATGT AACCCACTCG | 7560 |
| TGCACCAAC TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTCTGGGT GAGCAAAAAC | 7620 |
| AGGAAGGCAA AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT | 7680 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| ACTCTTCCTT | TTTCAATATT | ATTGAAGCAT | TTATCAGGGT | TATTGTCTCA | TGAGCGGATA | 7740 |
| CATATTTGAA | TGTATTTAGA | AAAATAAACA | AATAGGGGTT | CCGCGCACAT | TTCCCCGAAA | 7800 |
| AGTGCCACCT | GACGTCTAAG | AAACCATTAT | TATCATGACA | TTAACCTATA | AAAATAGGCG | 7860 |
| TATCACGAGG | CCCTTTCGTC | TTCAA | | | | 7885 |

**THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE
PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

- 1) A method of producing a trans-somatic mammal, wherein said method provides the incorporation of a DNA sequence into the secretory cells of the mammary gland to alter the composition of the milk, wherein said method comprising the steps of:
 - a) providing a vector containing a DNA sequence encoding a valuable compound;
 - b) packaging said vector into a cell line;
 - c) preparing a solution comprising the packaged vector and cell line producing said packaged vector; and
 - d) delivering said solution into the mammary gland to allow the incorporation of the DNA into the secretory cells of the mammary gland.
- 2) The method of Claim 1, wherein the method further comprises the step of flushing the mammary gland with an osmotically-balanced solution prior to delivering the said solution into the mammary gland.
- 3) The method of Claim 2, wherein the method further comprises the step of externally massaging the mammary gland several times a day after delivering the said solution into the mammary gland.
- 4) The method of Claim 2, wherein the method further comprises the step of growing the cells producing said packaged vector on a solid support means, and the solution comprises the cells on said support means and the packaged vector.
- 5) The method of Claim 4, wherein the method further comprises the step of externally massaging the mammary gland several times a day after delivering the said solution into the mammary gland.
- 6) The method of Claim 2 or 4, wherein a substance, with a density higher than the density

of the solution, is delivered into the mammary gland after the delivery of the solution, wherein said substance displaces the solution upwards in the mammary gland.

- 7) The method of Claim 1, wherein the valuable compound is a pharmaceutical.
- 8) The method of Claim 7, wherein the pharmaceutical is a compound selected from the group consisting of: a tissue plasminogen activator, an antibody, an antibiotic, a blood clotting factor, galactosyltransferase, a growth factor, an oncoprotein, a hormone, a milk protein, a hormone receptor, a tumor suppressor protein, a vaccine and an erythropoietin.
- 9) The method of Claim 8, wherein the pharmaceutical is a tissue plasminogen activator.
- 10) The method of Claim 1, wherein the vector is transiently transfected into PA317 cells; the resulting particles are harvested and trans-infected into PG13 cells.
- 11) The method of Claim 1, wherein the vector is selected from the group pL(X)SH, pL(X)SN, pLNS(X), pLHS(X), pLNC(X), pLHC(X), pLNA(X) and pLHA(X); wherein "A" is the beta actin promoter, "L" is the moloney murine virus long terminal repeat (LTR), "S" is the SV40 promoter, "C" is the cytomegalovirus promoter and "X" is a DNA sequence encoding a valuable compound.
- 12) The method of Claim 1, wherein the vector is selected from the group consisting of pL(X)iN and pL(X)iH; wherein "I" is an internal ribosomal entry site (IRES) and "X" is a DNA sequence encoding a valuable compound.
- 13) The method of Claim 1, wherein the vector is pLNM(X); wherein "M" is a mouse mammary tumor virus promoter and "X" is a DNA sequence encoding a valuable compound.
- 14) The method of Claim 1, wherein the vector is pLNM_i₂(X); wherein "M" is a mouse mammary tumor virus promoter, "i₂" is a modified wild type internal ribosomal entry site and "X" is a DNA sequence encoding a valuable compound.

- 15) The method of Claims 11 to 14, wherein "X" is a DNA sequence encoding a tissue plasminogen activator.
- 16) The method of Claim 1, wherein the solution is an aqueous solution.
- 17) The method of Claim 2, wherein the osmotically-balanced solution is a saline solution.
- 18) The method of Claim 4, wherein the solid support means is a matrix selected from the group consisting of Cytodex beads or Cultisphere.
- 19) The method of Claim 6, wherein said substance is a silicone substance.
- 20) The method of Claim 1, wherein the vector is transfected into a packaging cell line producing a non-retroviral derived particle.
- 21) The method of Claim 1, wherein the vector is transfected into a packaging cell line producing a retroviral derived particle.
- 22) The method of Claim 1, wherein the trans-infecting particle is produced *in vitro*.
- 23) The method of Claim 1, wherein the trans-infecting particle is produced *in vivo*.

1/2

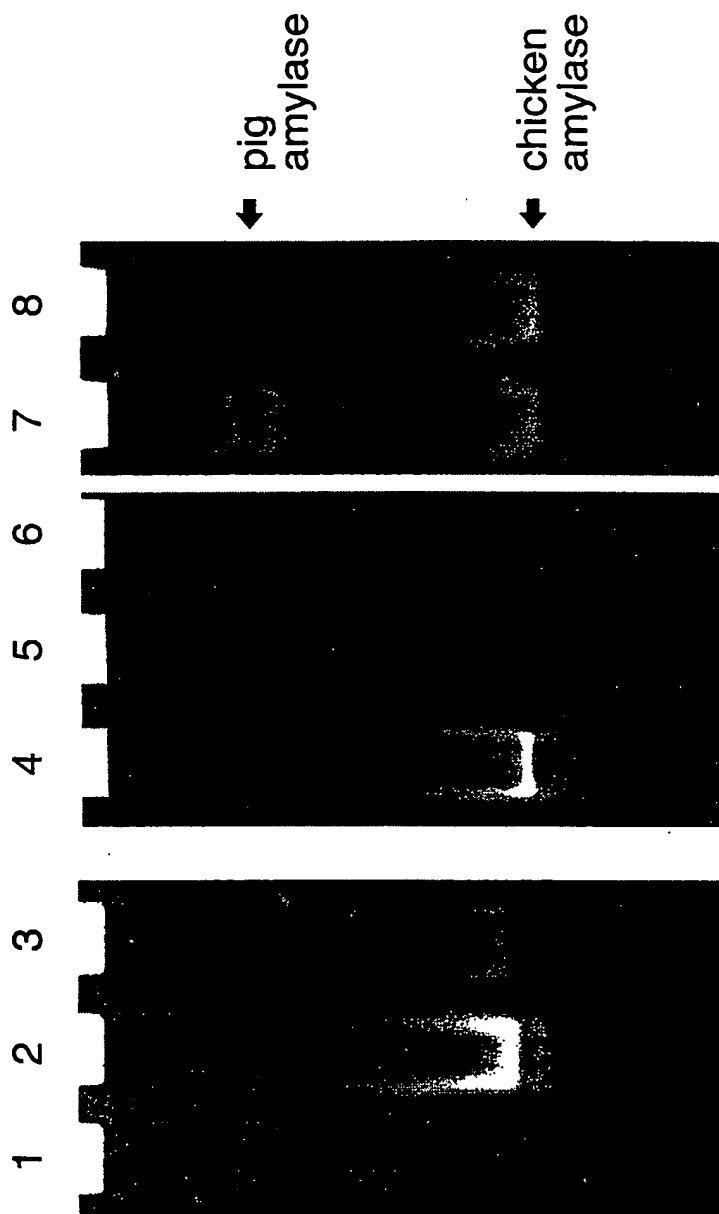


FIG. 1

2/2

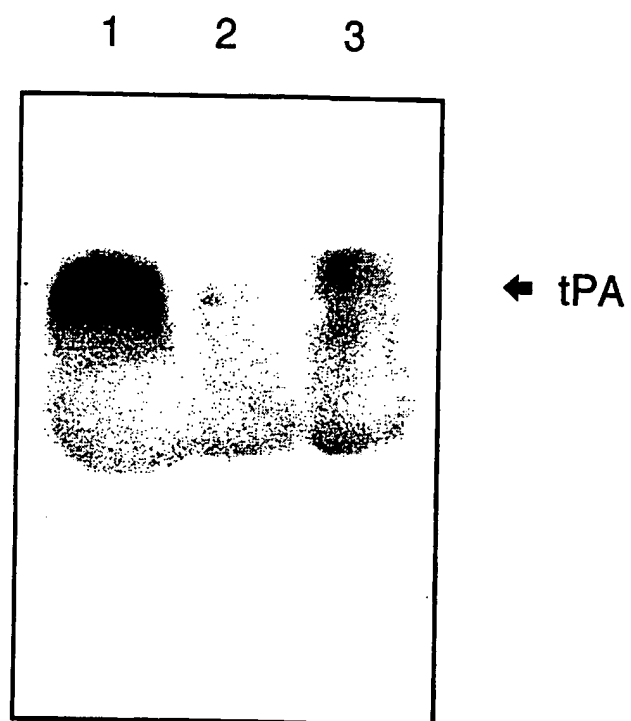


FIG. 2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 98/00607

| A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/00 A01K67/027 C12N15/86 C12N9/72 | | |
|---|--|--|
| According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 A01K C12N | | |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched | | |
| Electronic data base consulted during the international search (name of data base and, where practical, search terms used) | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | ARCHER, J.S. ET AL.: "Human growth hormone (hGH) secretion in milk of goats after direct transfer of the hGH gene into the mammary gland by using replication-defective retrovirus vectors" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 91, no. 15, 19 July 1994, pages 6840-6844, XP002088503 WASHINGTON US cited in the application | 1,7-11 |
| Y | see page 6844, column 1, line 12 - line 14 --- -/-- | 1-11 |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. | | |
| * Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "A" document member of the same patent family | | |
| Date of the actual completion of the international search 18 December 1998 | | Date of mailing of the international search report 13/01/1999 |
| Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fax: (+31-70) 340-3016 | | Authorized officer Chambonnet, F |

INTERNATIONAL SEARCH REPORT

International Application No

PCT/CA 98/00607

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|----------|---|-----------------------|
| Y | <p>YANG, N.S. ET AL.: "Gene transfer into mammalian somatic cells in vivo"</p> <p>CRITICAL REVIEWS IN BIOTECHNOLOGY, vol. 12, no. 4, 1992, pages 335-356, XP002088504</p> <p>see page 338, column 2</p> <p>see page 341, paragraph 3.</p> <p>---</p> | 1-11 |
| A | <p>WO 96 22379 A (NEXIA BIOTECHNOLOGIES INC)</p> <p>25 July 1996</p> <p>see page 24, line 8 - line 11; claims</p> <p>---</p> | 5,7-9,16 |
| A | <p>US 5 215 904 A (GOULD & WANG) 1 June 1993</p> <p>cited in the application</p> <p>see the whole document</p> <p>---</p> | 1 |
| Y | <p>CULVER, K.W. ET AL.: "In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors"</p> <p>SCIENCE., vol. 256, 12 June 1992, pages 1550-1552, XP002088505</p> <p>LANCASTER, PA US</p> <p>see the whole document</p> <p>---</p> | 1-11 |
| Y | <p>WO 95 15167 A (GENETIC THERAPY INC)</p> <p>8 June 1995</p> <p>see the whole document</p> <p>-----</p> | 1-11 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 98/00607

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|--|--|
| WO 9622379 A | 25-07-1996 | US 5780009 A AU 4428496 A CA 2210897 A EP 0805869 A | 14-07-1998 07-08-1996 25-07-1996 12-11-1997 |
| US 5215904 A | 01-06-1993 | NONE | |
| WO 9515167 A | 08-06-1995 | CA 2177998 A EP 0735887 A JP 9507839 T | 08-06-1995 09-10-1996 12-08-1997 |